VOL. 41 1950

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

EDITORIAL BOARD

Stephen M. Corey
Teachers College, Columbia University

J. B. Stroud
State University of Iowa

Jack W. Dunlap
10 E. 40th St., New York 17, N. Y.

Percival M. Symonds
Teachers College, Columbia University

Karl J. Holzinger
University of Chicago

Miles A. Tinker
University of Minnesota

Harold E. Jones
University of California

Alexander G. Wesman
The Psychological Corporation

H. H. Remmers
Purdue University

Paul A. Witty
Northwestern University

H. E. Buchholz
Managing Editor

Baltimore, Md.
WARWICK and YORK, Inc.
1950

Intergroup Attitudes of Gentile, Jewish and Apache Indian Children. Rose Zeligs 243-248

Mental Ability Ratings of Honor Students. Lillian G. Portenier 493-499

Note on the Evaluation of College Remedial Reading Courses, A. H. S. Robinson 83-96

Note on the Validity of the Mental Age Concept, A. Rukmini S. Ramaseshan 56-58

Notes on the Measurement of Mental Speed. Merle W. Tate 219-236

Origin and Development of the Spanish Attitude toward the Anglo and the Anglo Attitude toward the Spanish, The. Granville B. Johnson, Jr. . . . 428-439

Physical Maturing among Boys as Related to Behavior. Mary Cover Jones and Nancy Bayley 129-148

Predicting Success of Engineering Students. Ralph F. Berdie and Nancy A. Sutter 184-190

Psychology of Group Behavior; The Class as a Group. Wm. Clark Trow, Alvin E. Zander, William C. Morse, and David H. Jenkins 322-338

Publications Received 254-256, 380-384

Qualitative Differences in the Vocabulary Responses of Children. Herman Feifel and Irving Lorge . . 1-18

Relation between Intelligence and Achievement Test Results for a Group of Communities, The. Roger T. Lennon 301-308

Relationship of Reading and Speech Difficulties, The. Thomas H. Eames 51-55

Relative Effectiveness of Massed Versus Spaced Film Presentation, The. Philip Ash 19-30

Reliability and Validity of Involuntary Blinking as a Measure of Ease of Seeing. Miles A. Tinker . . 417-427

Reliability, Homogeneity and Number of Choices. N. L. GAGE AND DORA E. DAMRIN 385–404

Carl E. Seashore. 1866–1949. DANIEL STARCH . . . 217–218

Studies of the Group Rorschach in Relation to Success in the College of the University of Chicago. LEE J. CRONBACH 65–82

Study of High-school Academic Indices as a Criterion for College Admission, A. ISMAEL RODRIGUEZ BOU AND FRANKLIN L. STOVALL. 309–320

Study of Individual Differences in the Education of Teachers, The. FRANK S. FREEMAN 366–372

Study in Prognosis: The Guidance Value of Selected Measures of Musical Aptitude, Intelligence, Persistence, and Achievement in Tonette and Adaption Classes for Prospective Instrumental Students, A. HAROLD CARL MANOR. 31–50

Study of Rationalization, A. ROBERT COBB MYERS. . 149–160

Study of Teaching Potentialities, A. J. DAVID O'DEA 473–480

Technique for Evaluating the Ability of Teachers to Apply Principles Concerned with the Developmental Needs of Adolescent Girls. SARA ANN BROWN 481–487

Traits of Personality and Their Intercorrelations as Shown in Biographies. EDWARD L. THORNDIKE . 193–216

Truancy and Classroom Disorder as Symptoms of Personality Problems. FRANCES A. MULLEN 97–109

Use of Literal Grades, The. TRUMAN L. KELLEY. . . 488–492

Use of the Minnesota Multiphasic Personality Inventory in Screening College Students for Counseling Purposes. GLENN R. HAWKES. 116–121

Verbal Intelligence and Effectiveness of Participation in Group Discussion. NORMAN E. GREEN. . . . 440–445

What Educational Measurement in the Education of Teachers? WALTER W. COOK 339-347

What Teachers Should Know about the Psychology of Adolescence? GLENN M. BLAIR 356-361

What the Psychology of Learning Has to Contribute to the Education of the Teacher. G. LESTER ANDERSON 362-365

BOOK REVIEWS

Herschell Alt (Chairman) and Others. *Children Absent from School.* (MILES A. TINKER) 377

Lawrence A. Averill. *The Psychology of the Elementary-school Child.* (J. B. STROUD) 59

Susan Deri. *Introduction to the Szondi Test.* (GEORGE K. BENNETT) 117

Roy M. Dorcus and Margaret Hubbard Jones. *Handbook of Employee Selection.* (C. M. LOUTTIT) 502

Florence L. Goodenough. *Mental Testing: Its History, Principles, and Applications.* (LEE J. CRONBACH) . . . 122

Palmer O. Johnson. *Statistical Methods in Research* (JEROME E. DOPPELT) 249

Harold E. Jones. *Motor Performance and Growth.* (MILES A. TINKER) 373

Carney Landis and M. Marjorie Bolles. *Textbook of Abnormal Psychology.* (H. MELTZER) 500

Paul F. Lazarsfeld, and Frank N. Stanton, Editors. *Communications Research 1948-1949.* (N. L. GAGE) . . . 251

Charles Morris. *The Open Self.* (H. MELTZER) 61

Frederick Mosteller, Herbert Hyman, Phillip J. McCarthy, Eli S. Marks, and David B. Truman with the Collaboration of Leonard W. Doob, Duncan MacRae, Jr., Frederick F. Stephan, Samuel A. Stouffer and S. S. Wilks. *The Pre-election Polls of 1948: Report to the Committee on Analysis of Pre-election Polls and Forecasts.* (ROBERT L. JONES) 378

G. W. Parkyn. *Children of High Intelligence.* (GWEN F. ARNOLD) 60

Stanley D. Porteus. *The Porteus Maze Test and Intelligence.* (LEE J. CRONBACH) 502

T. W. Richards. *Modern Clinical Psychology.* (H. MELTZER) 127

Jean Paul Sartre. *The Psychology of Imagination.* (H. MELTZER) 62

Jean Paul Sartre. *The Emotions. Outline of a Theory.* (H. MELTZER) 125

Helen Shacter. *How Personalities Grow.* (MILES A. TINKER) 191

D. Snygg and A. W. Combs. *Individual Behavior.* (J. B. STROUD) 124

Alfred A. Strauss and Laura E. Lehtinen. *Psychopathology and Education of the Brain-injured Child.* (H. MELTZER) 123

Hubert Sorenson. *Psychology in Education.* (J. B. STROUD) 191

Willard L. Valentine and Delos D. Wickens. *Experimental Foundations of General Psychology.* (MILES A. TINKER) 446

J. E. Wallace Wallin. *Children with Mental and Physical Handicaps.* (H. MELTZER) 503

Ernest Glen Wever. *Theory of Hearing.* (LOUIS CHESLOCK) 374

Hugh Woodworth. *The Nature and Technique of Understanding: Some Fundamentals of Semantics.* (PHILIP M. KITAY) 374

INDEX TO AUTHORS

Anderson, G. Lester 362
Ash, Philip 19
Bayley, Nancy 129
Berdie, Ralph F 184
Blair, Glenn M 356
Bou, Ismael Rodriguez 309
Brown, Sarah Ann 481
Bruce, William F 348
Cook, Walter W 339
Cronbach, Lee J 65
Damrin, Dora E 385
Davenport, K. S 110
Deemer, W. L., Jr 173
Eames, Thomas H 51
Feifel, Herman 1
Freeman, Frank S 366
Fruchter, Benjamin 279
Gage, N. L 385
Grace, Harry A 161
Green, Norman E 440
Grumby, Mary H 405
Hawkes, Glenn R 116
Jenkins, David H 322
Johnson, Granville B, Jr 257, 428
Jones, Mary Cover 120
Kelley, Truman L 488
Klausmeier, Herbert 449
Lennon, Roger T 301
Lorge, Irving 1
Manion, Harold Carl 31
Morse, William C 322
Mullen, Frances A 97
Myers, Robert Cobb 149
O'Dea, J David 473
Perkins, Gladys L 405
Portenier, Lillian G 493
Rafferty, J. A 173
Ramaseshan, Rukmini S 56
Remmers, H. H 110
Robinson, H S 83
Schlesser, George E 237
Shuey, Audrey M 292
Starch, Daniel 217
Stovall, Franklin L 309
Sutter, Nancy A 184
Swanson, Donovan 449
Tate, Merle W 219
Thorndike, Edward L 193
Tinker, Miles A 417
Trow, Wm Clark 322
Zander, Alvin F 322
Zeligs, Rose 243

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 41 January, 1950 Number 1

QUALITATIVE DIFFERENCES IN THE VOCABULARY RESPONSES OF CHILDREN*

HERMAN FEIFEL
Adjutant General's Office†

and

IRVING LORGE
Teachers College, Columbia University

A. INTRODUCTION

The intimate relation of vocabulary with the thought processes of the individual has long been recognized. Terman,[14] discussing the tentative revision of the Binet-Simon scale in 1912, stated that a qualitative analysis of the verbatim definitions of children of different ages and mentality would be of great value. In 1916 Terman[15] found that children of six years, as a rule, defined *objects in terms of use, and that it was not before eight years that* two-thirds of the children spontaneously gave a large proportion of definitions in terms superior to use. He also emphasized the point that the type of definition used by the child, i.e., whether thought of in terms of use, appearance, material comprising it, or class relationship, threw interesting light on the maturity of his apperceptive processes, and that one could differentiate at least a half-dozen degrees of excellence in definition related to the mental level of the subject. He admitted that although the form of definition was important, it had not been taken into consideration in the scoring of the test. Piaget[11,12] also in his work with children showed that words are bound up with cognition and are related to concept formation.

---

* The authors wish to express their appreciation to Anne Dragositz of the Educational Testing Service, New York, and to Professor Nicholas Hobbs of Teachers College, Columbia University, for making available the raw data on which this study is based.

† The views herein stated do not necessarily reflect those of the Department of the Army.

In testing the intelligence of children by means of vocabulary, two main types of vocabulary tests have been employed. One kind, relying on recall factors, presents a key word to which the child must respond with an acceptable definition. The second variety, stressing recognition features, introduces a word and then four or five alternative words (multiple-choice technique) from which the child chooses the most appropriate matching one. The scoring system of the recall type assumes that all right answers are of equal value and makes no differentiation between a 'good' right answer and a 'poor' right answer, except for occasional half-credits, nor does it take into account the various shades, quality, and range of possible meanings.

The purpose of the present study aims at a further clarification and extension of our knowledge concerning the successive stages in concept formation and developmental thinking of children through a qualitative analysis of their verbatim responses to the Form L Vocabulary Test of the Revised Stanford-Binet Scale.

B. REVIEW OF LITERATURE

As far back as 1904 Chambers[3] made a significant study of growth in the general meaning of vocabulary words. He indicated that in the early years of life we have an accurate knowledge of only those things which are most immediate and familiar, and that in studying the development of the child's use of language the most fruitful field is found in an analysis of the quality and expansion of vocabulary during the school years. Kirkpatrick[8] stated that growth with age occurred not only in the vocabulary range but also in the character of the definitions attached to words. Binet and Simon,[2] studying the development of intelligence in children, discovered that children of age six defined words in terms of use and that it was not until they reached age nine that most of the definitions they gave were superior to use. Terman[14,15] has long emphasized the importance of qualitative responses in vocabulary and the need for their analysis. Marx,[9] carrying out a qualitative study of the first fifty words of the 1916 Stanford-Binet vocabulary on a large group of children and adults, reported that the highest quality types of definition in relation to chronological age were those of the synonym and genus variety. Lowest on the age scale were those definitions using illustration or example, use, and the repetition types of response. Green[6] also qualitatively analyzed the responses of school children and adults on fifty vocabulary words (forty-five of which later became the Form L Vocabulary Test of the 1937 Revised Stanford-Binet). She noted that ages six and seven were characterized by use definitions, and that types of definition having lower median ages than use were repetition and demonstration. Gray and Holmes[5] emphasized that the character of the definitions attached to words changed notably from the lower to higher grades, and that the period from nine to fifteen years of age was particularly productive in the acquisition of different types of meaning. Reichard and Rapaport,[13] in a study of the responses of children and adolescents to the Similarities and Differences Test and other conceptual items of the Stanford-Binet Scale, found that 'concrete' definitions decreased with age and gave place to 'functional' and then to 'conceptual' types of definition. Wechsler[17] has stated that in defining a word a subject gave more than just its mere meaning. From the clinical point of view, the character and quality of the word definition given by the individual permitted insight into his thought processes.

Analyses of the qualitative studies of vocabulary stress the fact that growth with age occurs not only in the vocabulary range but in the character of the word definition as well, and that the quality and completeness of word definition change considerably from the lower to higher grades. It indicates that there are many possible ways of defining a word, some of which are easy, others comparatively difficult. Both a six year old and a fourteen year old know the meaning of the word 'straw.' The six year old defines it as 'straw hat,' the fourteen year old as 'dried hay.' Obviously there are characteristic differences in thinking involved, and yet, in both cases, our present system of scoring the vocabulary test gives similar credit to the responses.

C. METHOD

1. *Selection of the Test.* The Form L Stanford-Binet Vocabulary Test was chosen because it extends from the six-year level to the adult level, and the definitions given to its forty-five words allow qualitative differences in the responses to show themselves. In addition, the Test has excellent standardization, good interest value, and presents a familiar task to the child.

However, one should bear in mind that not all of the words of the Test permit a full range of qualitative differences to express themselves in the verbatim responses. This is particularly true for the more difficult words in the list. Whereas for a word like 'orange' one can reply with various types of acceptable answers, e.g., 'color' (synonym), 'citrus fruit' (synonym modified), 'you eat it' (use), 'it's round' (description), etc., the correct answers for a word like 'piscatorial' are usually limited to a synonym or synonym-type of definition, e.g., 'pertaining to fishes' or 'fishlike.'

2. *Qualitative Classification of Responses.* On the basis of Feifel's[4] work in classifying the variety of different qualitative responses given to the words of the Form L Vocabulary Test a fivefold qualitative classification system was set up for each definition of every word. One category consisted of synonym types of response. Another included use and description types of definition. These were combined into one category because they were found to occur at approximately the same developmental level in children. A third category contained the explanation type of response. A fourth one was made up of the demonstration, illustration, inferior explanation, and repetition types of response. These were included together because empirical analysis also indicated that they occurred at about the same developmental level. The final category was composed of all types of error response. These categories were kept uniform for the entire list of forty-five words. Here are examples for each of the five categories:

Synonym Category

(a) Synonym unmodified: Orange = a fruit
(b) Synonym modified by use: Straw = hay that cattle eat
(c) Synonym modified by description: Gown = long dress
(d) Synonym modified by use and description: Eyelash = hair over the eye that protects you
(e) Synonym qualified as to degree: Tap = touch lightly

Use, Description, and Use and Description Category

(a) Use: Orange = you eat it
(b) Description: Straw = it's yellow
(c) Use and Description: Orange = you eat it and it's round

Explanation Category

(a) Explanation: Priceless = it's worth a lot of money
Skill = being able to do something well

Demonstration, Repetition, Illustration and Inferior Explanation Category

(a) Demonstration:	For words like tap, eyelash, etc.
(b) Repetition:	Puddle = a puddle of water
(c) Illustration:	Priceless = a gem
(d) Inferior Explanation:	Scorch = hot

Error Category

(Incorrect Demonstration, Misinterpretation, Wrong Definition, Clang Association, Repetition without Explanation, Omits)

(a) Incorrect Demonstration:	Eyelash = points to eyebrow
(b) Misinterpretation:	Regard = protects something
(c) Wrong Definition:	Orange = a vegetable
(d) Clang Association:	Roar = raw; skill = skillet
(e) Repetition without Explanation:	Puddle = puddle
(f) Omits:	When the word is left out

In all the cases used in this study the Vocabulary Test was first administered and scored according to Terman's[16] directions by experienced examiners. After checking the quantitative scoring, all of the verbatim word definitions were rescored in terms of the five qualitative categories just outlined.

3. *Reliability of Scoring.*—In measuring the reliability of qualitative categories a common procedure is to compare item by item, or response by response, the records obtained when two or more independent workers record the same behavior, or score the same responses. Agreement can be computed in terms of per cents. To establish the reliability of the qualitative scoring, fifty of the cases were rescored by an experienced examiner not specially trained in the qualitative method of scoring. Employing the Arrington[1] formula where the responses in each observer's scoring that agree with the other's (in effect, doubling the agreements) is divided by this total plus the disagreements (responses dissimilarly recorded and responses noted by one observer and omitted by the other,* i.e., $\frac{2 \times \text{agreements}}{2 \times \text{agreements} + \text{disagreements}}$, the range of the per cent of agreements for the forty-five words of the Test was found to extend from 97 to 100 per cent. They are of such an order to signify that the qualitative scoring was carried out with a high degree of consistency.

---

* There were no responses in any of the fifty cases noted by one observer and not by the other. This part applies essentially to workers recording observational types of behavior.

4. *Subjects.*—The tested population consisted of nine hundred school children, male and female, ranging in age from six through fourteen years, with one hundred children, practically equally divided as to sex, at each year level. The data were secured from testing carried out in a survey of public schools in the Eastern seaboard states during the years 1939 to 1942, and from testing carried out in advanced psychological testing courses at Teachers College, Columbia University during 1946 and 1947. All the children were of the white race, American born, and gave no indication of any language or physical handicaps. Whenever any doubt arose as to the validity of a test result, the case was not used.

In the first tabulations, results for both sexes were kept separate. Since analyses showed no significant differences but rather similar pictures, the data for the males and females were combined and only total group determinations, in this respect, are reported in the present study.

Educational level was determined by either completion of or attendance in a grade. The years of schooling ranged from one to ten years.

Table 1 presents the background data of the nine hundred children used in this study. The children are slightly above average in intelligence at each year level except at age fourteen. They were matched against the composite Form L-M age-grade distribution and for mean IQ by age and grade distribution as used by McNemar[12] in the standardization of the Revised Stanford-Binet Scale. The ages of the children ranged from six through fourteen, the grades from I through X. In his sample of 1623 for both Form L and Form M, McNemar had mean IQ's of 103.2 and 101.0 and SD's of 16.8 and 16.6, respectively. McNemar indicated that corrections for the inadequate rural representation and bias in the occupational status of the parents would bring the IQ means much closer to 100. The data of this study, too, are biased insofar as there is little rural representation.

D. TREATMENT OF THE DATA

Table 2 presents the means and standard deviations for each of the five qualitative categories at each age level. The data demonstrate that as the children get older they use the synonym types of response more and more often, and that at age nine and

TABLE 1.—MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS FOR MA, IQ, EDUCATION, AND VOCABULARY SCORE CORRECT BY AGE CLASSIFICATIONS (N = 900)

	MA	IQ	Educ	Vocab. Score Correct	MA	IQ	Educ	Vocab. Score Correct	MA	IQ	Educ	Vocab. Score Correct
	6 Years (N = 100)				7 Years (N = 100)				8 Years (N = 100)			
M	80.1*	103.3	1.0	6.2	91.1	102.3	1.6	7.3	104.2	103.0	2.3	8.3
SD	9.5*	10.9	.10	1.6	13.0	13.0	.51	2.4	15.0	14.8	.55	2.6
$r_{(MA)}$		.93	.06	.55		.96	.05	.73		.97	.21	.76
$r_{(IQ)}$			.04	.44			.05	.71			.09	.75
$r_{(Educ)}$				.05				.37				.28
	9 Years (N = 100)				10 Years (N = 100)				11 Years (N = 100)			
M	117.4	104.1	3.5	9.9	128.9	103.3	4.2	11.9	141.1	103.1	5.2	13.8
SD	17.7	15.7	.65	3.3	19.4	15.4	.73	3.9	23.8	17.4	.84	4.7
$r_{(MA)}$		.98	.34	.77		.98	.50	.78		.99	.64	.86
$r_{(IQ)}$			.33	.79			.50	.77			.62	.84
$r_{(Educ)}$				.32				.48				.61
	12 Years (N = 100)				13 Years (N = 100)				14 Years (N = 100)			
M	152.0	102.3	6.0	15.4	165.7	103.8	7.1	17.6	167.2	100.6	8.0	17.4
SD	22.7	15.0	1.0	4.5	23.4	14.6	.94	4.1	22.8	13.4	1.0	4.5
$r_{(MA)}$		.99	.73	.84		.99	.64	.80		.96	.63	.73
$r_{(IQ)}$			.73	.84			.62	.79			.61	.74
$r_{(Educ)}$				.68				.54				.59

* In months

TABLE 2.—MEANS AND STANDARD DEVIATIONS OF FIVE QUALITATIVE CATEGORIES BY AGE CLASSIFICATION

Age Group	N	Synonym		Use and Description		Explanation		Demonstration, Illustration, etc.		Error	
		M	SD	M	SD	M	SD	M	SD	M	SD
6	100	1.0	1.3	3.4	1.4	.6	.7	1.2	1.2	38.8	1.6
7	100	1.9	1.8	2.9	1.4	.8	.8	1.6	1.2	37.7	2.5
8	100	2.7	2.2	2.7	1.4	1.1	1.1	1.8	1.4	36.7	2.6
9	100	3.7	2.9	2.6	1.5	1.4	1.3	2.1	1.4	35.1	3.3
10	100	5.5	3.8	2.4	1.5	2.1	1.8	2.1	1.4	33.0	3.9
11	100	7.3	4.8	2.1	1.3	2.4	1.7	1.9	1.4	31.3	4.8
12	100	8.6	4.6	2.1	1.2	3.0	1.8	1.6	1.4	29.7	4.7
13	100	10.6	5.0	1.8	1.5	3.6	1.9	1.6	1.4	27.4	4.1
14	100	11.6	5.1	1.8	1.4	3.2	1.9	1.4	1.4	27.6	4.6

higher they employ them more frequently than any other type of definition. The use and description types of response are given more often than any other type of definition by the children at ages six and seven. The older children, however, select them less frequently than any other type of response except the illustration, demonstration, inferior explanation, and repetition category types of response. The children use the explanation type of definition least of all at ages six through nine years, but as they get older choose it more often than all other kinds of response except the synonym variety. The demonstration, repetition, illustration, and inferior explanation types of response do not appear to be given frequently at any age. The nine- and ten-year olds give them more often than do any of the other age groups. From age eleven and up this category of response is employed less often than any other category. The error category, as was to be expected, shows a steady decline in frequency of choice as the children get older.

A graphic presentation of the data is given in Figure 1. It is apparent that at ages six and seven the children give the use and description types of response most often, and the explanation type of definition least often. At ages nine and ten, the synonym types of response are used most frequently by the children, the

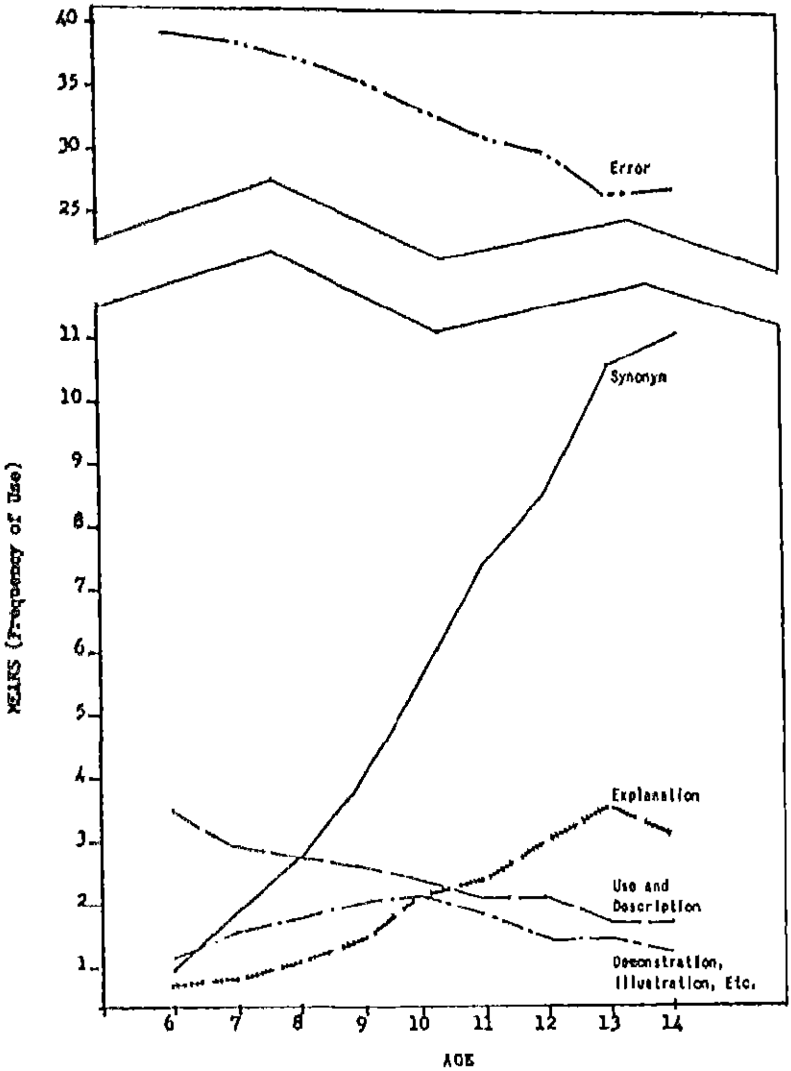


Fig. 1. Mean frequency of use of five qualitative categories by age.

use and description types of definition are given less frequently than at ages six and seven, and the explanation type of response is given increasingly more often. From age eleven and up, the children very definitely use the synonym types of response most

TABLE 3.—SIGNIFICANCE OF THE DIFFERENCES BETWEEN YOUNGER AND OLDER CHILDREN BY AGE GROUPS ON SYNONYM TYPES OF RESPONSE (N = 100 FOR EACH AGE GROUP)

Age Group		6	7	8	9	10	11	12	13
7	D†	−.9							
	$\frac{D}{SE_D}$	−4.3*							
8	D	−1.7	−.8						
	$\frac{D}{SE_D}$	−6.7*	−2.9*						
9	D	−2.7	−1.8	−1.0					
	$\frac{D}{SE_D}$	−8.6*	−5.5*	−2.8*					
10	D	−4.5	−3.6	−2.8	−1.8				
	$\frac{D}{SE_D}$	−11.2*	−8.6*	−6.2*	−3.7*				
11	D	−6.3	−5.4	−4.0	−3.6	−1.8			
	$\frac{D}{SE_D}$	−12.6*	−10.5*	−8.6*	−6.4*	−3.0*			
12	D	−7.6	−6.7	−5.9	−4.9	−3.1	−1.3		
	$\frac{D}{SE_D}$	−15.5*	−13.3*	−11.2*	−8.8*	−5.2*	1.9		
13	D	−9.6	−8.7	−7.9	−6.9	−5.1	−3.3	−2.0	
	$\frac{D}{SE_D}$	−18.5*	−16.3*	−14.3*	−12.0*	−8.3*	−4.8*	−2.9*	
14	D	−10.0	−9.1	−8.3	−7.3	−5.5	−3.7	−2.4	−.4
	$\frac{D}{SE_D}$	−18.9*	−16.8*	−14.8*	−12.6*	−8.8*	−5.3*	−3.5*	−.6

D† = Difference between obtained means for age groups 6 and 7
* = Significant at .01 level.
− = Minus values indicate that the difference is in favor of older age group.

often. They are followed in frequency of use by the explanation type of definition, the use and description kinds of response, and finally, the demonstration, illustration, inferior explanation, and repetition types of response.

1. *Significant Differences in Qualitative Responses Between*

TABLE 4.—SIGNIFICANCE OF THE DIFFERENCES BETWEEN YOUNGER AND OLDER CHILDREN BY AGE GROUPS ON USE AND DESCRIPTION TYPES OF RESPONSE ($N = 100$ FOR EACH AGE GROUP)

Age Group		6	7	8	9	10	11	12	13
7	D	.5							
	$\frac{D}{SE_D}$	2.5*							
8	D	.7	.2						
	$\frac{D}{SE_D}$	3.6†	1.1						
9	D	.8	.3	.1					
	$\frac{D}{SE_D}$	4.0†	1.6	.4					
10	D	1.0	.5	.3	.2				
	$\frac{D}{SE_D}$	5.1†	2.6†	1.5	1.2				
11	D	1.3	.8	.6	.5	.3			
	$\frac{D}{SE_D}$	6.7†	4.2†	3.1†	2.8†	1.0			
12	D	1.3	.8	.6	.5	.3	0		
	$\frac{D}{SE_D}$	7.1†	4.5†	3.3†	3.0†	1.7	0		
13	D	1.6	1.1	.9	.8	.6	.3	.3	
	$\frac{D}{SE_D}$	8.0†	5.5†	4.4†	4.0†	2.0†	1.3	1.3	
14	D	1.6	1.1	.9	.8	.6	.3	.3	0
	$\frac{D}{SE_D}$	7.8†	5.3†	4.2†	3.9†	2.7†	1.1	1.1	0

* = Significant at .05 level.
† = Significant at .01 level.

Younger and Older Children.—In order to discover whether there were any significant differences between the type of qualitative definitions given by the younger children as against the older ones to the forty-five words of the Vocabulary Test, the mean differences were determined for each category at all age levels and an evaluation made of the significance of these differences in terms of the standard error of their differences. The results appear in Tables 3, 4, 5, 6, and 7.

Table 3 presents the results for the synonym types of response

TABLE 5.—SIGNIFICANCE OF THE DIFFERENCES BETWEEN YOUNGER AND OLDER CHILDREN BY AGE GROUPS ON EXPLANATION TYPE OF RESPONSE (N = 100 FOR EACH AGE GROUP)

Age Group		6	7	8	9	10	11	12	13
7	D	−.2							
	D/SE_D	−1.8							
8	D	−.5	−.3						
	D/SE_D	−4.3†	−2.4*						
9	D	−.8	−.6	−.3					
	D/SE_D	−5.1†	−3.6†	−1.8					
10	D	−1.5	−1.3	−1.0	−.7				
	D/SE_D	−7.5†	−6.5†	−4.6†	−2.8†				
11	D	−1.8	−1.6	−1.3	−1.0	−.3			
	D/SE_D	−9.4†	−8.3†	−6.4†	−4.4†	−1.5			
12	D	−2.4	−2.2	−1.9	−1.6	−.9	−.6		
	D/SE_D	−12.7†	−11.7†	−9.6†	−7.3†	−4.1†	−2.7†		
13	D	−3.0	−2.8	−2.5	−2.2	−1.5	−1.2	−.6	
	D/SE_D	−14.0†	−12.4†	−11.0†	−8.8†	−5.8†	−4.4†	−1.99*	
14	D	−2.6	−2.4	−2.1	−1.8	−1.1	−.8	−.2	.4
	D/SE_D	−12.2†	−10.8†	−9.4†	−7.3†	−4.4†	−3.0†	−.6	1.3

* = Significant at .05 level.
† = Significant at .01 level.
− = Minus values indicate that the difference is in favor of older age group.

and indicates that the older children significantly more often employ this type of definition than do the younger children. This holds true for each succeeding year level from age six and up except between the age levels of eleven and twelve, and thirteen and fourteen.

The data for the use and description types of response appear in Table 4. They show that the younger children, age six through nine years, significantly more often give these types of

TABLE 6.—SIGNIFICANCE OF THE DIFFERENCES BETWEEN YOUNGER AND OLDER CHILDREN BY AGE GROUPS ON DEMONSTRATION, ILLUSTRATION, INFERIOR EXPLANATION, AND REPETITION TYPES OF RESPONSE (N = 100 FOR EACH AGE GROUP)

Age	Group	6	7	8	9	10	11	12	13
7	D	−.4							
	$\frac{D}{SE_D}$	−2.9†							
8	D	−.6	−.2						
	$\frac{D}{SE_D}$	−3.5†	−1.1						
9	D	−.9	−.5	−.3					
	$\frac{D}{SE_D}$	−5.4†	−3.1†	−1.7					
10	D	−.9	−.5	−.3	0				
	$\frac{D}{SE_D}$	−5.2†	−2.8†	−1.5	0				
11	D	−.7	−.3	−.1	.2	.2			
	$\frac{D}{SE_D}$	−4.2†	−1.90	−.7	1.0	.8			
12	D	−.4	0	.2	.5	.5	.3		
	$\frac{D}{SE_D}$	−2.4*	0	1.0	2.6†	2.4*	1.6		
13	D	−.4	0	.2	.6	.5	.3	0	
	$\frac{D}{SE_D}$	−2.8†	0	.6	2.3*	2.1*	1.3	0	
14	D	−.2	.2	.4	.7	.7	.5	.2	.2
	$\frac{D}{SE_D}$	−.9	1.4	2.2*	3.8†	3.0†	2.8†	1.2	1.0

* = Significant at .05 level.
† = Significant at .01 level.
− = Minus values indicate that the difference is in favor of older age group.

definition than do the older children, age eleven through fourteen years. And although no significant differences exist among the older children, there is a general tendency for the thirteen- and fourteen-year olds to use these types of response less frequently than do the eleven- and twelve-year olds.

Table 5 illustrates the findings for the explanation type of response and indicates that the children age six through nine

TABLE 7.—SIGNIFICANCE OF THE DIFFERENCES BETWEEN YOUNGER AND OLDER CHILDREN BY AGE GROUPS ON ERROR TYPES OF RESPONSE (N = 100 FOR EACH AGE GROUP)

Age Group		6	7	8	9	10	11	12	13
7	D	1.1							
	$\frac{D}{SE_D}$	3.5†							
8	D	2.1	1.0						
	$\frac{D}{SE_D}$	6.7†	3.0†						
9	D	3.7	2.6	1.6					
	$\frac{D}{SE_D}$	10.3†	6.6†	3.8†					
10	D	5.8	4.7	3.7	2.1				
	$\frac{D}{SE_D}$	13.7†	10.3†	7.7†	4.2†				
11	D	7.5	6.4	5.4	3.8	1.7			
	$\frac{D}{SE_D}$	14.7†	11.0†	9.7†	6.0†	2.7†			
12	D	9.1	8.0	7.0	5.4	3.3	1.6		
	$\frac{D}{SE_D}$	18.1†	15.1†	12.8†	9.4†	5.4†	2.4*		
13	D	11.4	10.3	9.3	7.7	5.6	3.9	2.3	
	$\frac{D}{SE_D}$	25.4†	21.6†	18.0†	14.0†	9.0†	6.3†	3.8†	
14	D	11.2	10.1	9.1	7.5	5.4	3.7	2.1	−.2
	$\frac{D}{SE_D}$	22.0†	19.5†	17.1†	13.4†	9.1†	5.7†	3.2†	−.3

* = Significant at .05 level.
† = Significant at .01 level.
— = Minus values indicate that the difference is in favor of the older age group.

years employ this category of definition response significantly less frequently than do the older children, age ten through fourteen years. The twelve-, thirteen-, and fourteen-year olds seem to use this type of response about equally.

The results for the demonstration, illustration, inferior explanation, and repetition types of response appear in Table 6. They show that the six year old children give these types of definition significantly less often than do the older children except for the fourteen-year olds, and that the nine- and ten-year old children employ them significantly more often than do the six- and seven-year olds. However, the eight year old children use these types of response significantly more often than do the fourteen-year olds, as do the nine- and ten-year olds when compared with the twelve-, thirteen-, and fourteen-year olds, and the eleven-year olds when compared with the fourteen-year olds. It should also be recalled that these types of definition are given less frequently by the older children, age eleven through fourteen years, than any other qualitative category of response.

Table 7 presents the findings for the error types of definition and shows, as expected, that they are made significantly less often by the older children than by the younger children. These significant differences hold true for each succeeding year level from age six and up, except between the ages of thirteen and fourteen years.

To illustrate some of the differences found in the qualitative responses given to the words of the Vocabulary Test by the younger and older children, a few typical definitions are listed below to the following words of the Test, i.e., 'puddle,' 'tap,' and 'gown.' To the word 'puddle' the younger children responded with answers like the following: 'a puddle of water,' 'mud,' 'what you step in,' etc. The older children responded to the same word with definitions of the following kind: 'a small pool of water,' 'water that gathers after a rain,' etc. To the word 'tap' the younger children gave replies of the following variety: 'tap your foot,' 'tap on a door,' demonstrated the action with their fingers or foot, etc. The older children gave definitions of the following kind to the same word: 'a faucet,' 'a light knock,' 'kind of a rap,' etc. To the word 'gown' the younger children responded with definitions of the following type: 'you wear it,' 'nightgown,' 'what you sleep in,' etc. The older children responded with answers of the following kind: 'an evening dress,' 'a long dress that women wear,' 'a beautiful dress you put on going to dances,' etc.

It should be cautioned, however, that no particular type of definition response is the exclusive possession of any specific age group, or, for that matter, any group.

E. DISCUSSION

Significant differences have been demonstrated in the qualitative types of response given by younger children when compared with older children of similar background. Analysis of their verbatim responses shows that the older children significantly more often use the explanation and synonym types of definition, whereas the younger children significantly more often employ the use and description, and demonstration, illustration, inferior explanation, and repetition types of response. These findings substantiate the work of Binet and Simon[2] and Terman[15] who found that the use type of definition is most frequently given by children eight years of age and younger. In addition, they confirm Green's[6] work in showing that use, description, repetition and demonstration types of response are more characteristic of younger than of older children. The results also support the findings of Piaget[11,12] and Reichard and Rapaport[13] in evidencing changes in the conceptual level of development of children as they grow older. To the younger child an 'orange' is 'what you eat' or 'it's round.' Only as he grows older does he see it as a 'citrus fruit' or 'a fruit which grows in California or Florida.' Obviously, characteristic differences exist in the thinking of younger children when they are compared with older children.

The younger children tend to perceive words as concrete ideas and do not generalize. They emphasize the particular or isolated aspect rather than the categorical or 'class' feature of the meaning as older children do. A 'puddle' means 'mud' rather than a 'small pool of water,' and 'straw' is 'yellow' rather than 'dried grass.' This emphasis of younger children on the concrete approach as evinced through their word definitions with a personal rather than symbolic outlook shows striking similarity to the qualitative types of definition given by abnormals. Feifel,[4] employing the same fivefold qualitative system used in this study, demonstrated that in response to the words of the Form L Stanford-Binet Vocabulary Test abnormals, when compared with normals similar in background, significantly more often gave the use and description, and demonstration, illustration, inferior explanation, and repetition types of definition than they did the explanation and synonym types of response.

Since there is a definite relation between age and the types of qualitative definition children use in defining words, and since the method of approach used by individuals is as revealing as the mere passing or failing of a word, study of the qualitative patterns of response given by children to vocabulary words can throw light on the conceptual level or mode of thinking of the child. This may prove helpful in supporting and supplementing other data we possess concerning the child's thinking processes.

F. SUMMARY AND CONCLUSIONS

1) The Form L Stanford-Binet Vocabulary Test was administered to nine hundred children between the ages of six and fourteen, and qualitative analysis made of their verbatim responses by means of a fivefold qualitative category system.

2) Significant differences were established between the qualitative responses given by the younger children as against those used by the older children. The younger children significantly more often employed the use and description, and illustration, demonstration, inferior explanation, and repetition types of response, whereas the older children significantly more often used the synonym and explanation types of response.

3) Characteristic differences exist in the thinking of younger children when compared with older children similar in background.

4) Younger children perceive words as 'concrete' ideas and emphasize their isolated or particular aspects, whereas older children stress the abstract or 'class' features of the word meanings.

5) Strong similarity is disclosed between the qualitative responses given to the Form L Stanford-Binet Vocabulary Test by younger children and abnormals.

6) Qualitative analysis of the vocabulary responses of children can serve as an aid in supplementing other sources of data concerning the child's thinking processes.

BIBLIOGRAPHY

1) Ruth E. Arrington. *Interrelations in the Behavior of Young Children.* Child Development Monograph No. 8. New York: Teachers College, Columbia University, 1932.

2) A. Binet, and T. Simon. *The Development of Intelligence in*

Children. Publications of the Training School at Vineland, New Jersey, 1916, pp. 336.

3) W. G. Chambers. "How words get meaning." *Pedagog. Semin.*, 1904, 11, 30–50.

4) H. Feifel. "Qualitative differences in the vocabulary responses of normals and abnormals." *Genet. Psychol. Monogr.*, 1949, 39, 151–204.

5) W. S. Gray, and Eleanor Holmes. *The Development of Meaning Vocabularies in Reading.* Publications of the Laboratory Schools of The University of Chicago, No. 6, 1938, pp. 140.

6) Helen J. Green. *A Qualitative Method for Scoring the Vocabulary Test of the New Revision of the Stanford-Binet.* Unpublished Master's thesis, Stanford University, 1931.

7) J. P. Guilford. *Psychometric Methods.* New York: McGraw-Hill, 1936, pp. 566.

8) E. A. Kirkpatrick. "A vocabulary test." In W. S. Gray and Eleanor Holmes, *The Development of Meaning Vocabularies in Reading.* University of Chicago, 1938, p. 17.

9) Barbara Marx. *A Study of the First Fifty Words of the Stanford-Binet Vocabulary.* Unpublished Master's thesis, Stanford University, 1928.

10) Q. McNemar. *The Revision of the Stanford-Binet Scale.* Boston: Houghton Mifflin, 1942, pp. 185.

11) J. Piaget. *The Language and Thought of the Child.* New York: Harcourt, 1926, pp. 246.

12) ———. *Judgment and Reasoning in the Child.* New York: Harcourt, 1928, pp. 260.

13) Suzanne Reichard and David Rapaport. "The rôle of testing concept formation in clinical psychological work." *Bull. Menninger Clin.*, 1943, 3, 99–105.

14) L. M. Terman and H. G. Childs. "A tentative revision and extension of the Binet-Simon measuring scale of intelligence." *J. educ. Psychol.*, 1912, 3, 198–208.

15) L. M. Terman. *The Measurement of Intelligence.* Boston: Houghton Mifflin, 1916, pp. 362.

16) L. M. Terman and Maud A. Merrill. *Measuring Intelligence.* Boston: Houghton Mifflin, 1937, pp. 461.

17) D. Wechsler. *The Measurement of Adult Intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944, pp. 258.

THE RELATIVE EFFECTIVENESS OF MASSED VERSUS SPACED FILM PRESENTATION[1]

PHILIP ASH
The Pennsylvania State College

INTRODUCTION

While the average classroom lecture is usually between forty and sixty minutes long, and the average entertainment film lasts eighty or ninety minutes, the typical instructional film is designed to run exactly ten minutes. A wide variety of factors contribute to this situation. As a practical matter the instructional film producers have more or less tailored their films to fit the standard 400-foot reel for 16 millimeter projectors. Films are frequently, perhaps usually, designed as instructional aids, to assist rather than supplant the lectures. Beyond these reasons, however, is the often expressed feeling that longer films are ineffective teachers.

These limits may be adequate for ordinary school practice. In large mass training programs, however, such as those conducted by the Armed Forces, the question of how long film sessions can last and still be effective often arises. Lack of instructors and problems of scheduling frequently make it desirable that sessions much longer than ten or fifteen minutes be arranged.

STATEMENT OF THE PROBLEM

The experimental problem may be stated as follows: Given a body of information that has been translated into films, do people learn more if they are presented with this content in one long film in a single session, or if they are presented with the content broken up into several short units in two or more sessions?

---

[1] Abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at The Pennsylvania State College, June 1949. The research was made possible by a fellowship grant from The Instructional Film Research Program, The Pennsylvania State College, under contract with the Special Devices Center, Office of Naval Research, Navy Department, Contract No. N6onr-269 T.O. VII. Grateful acknowledgment is made herewith to the personnel of the Film Research Program who aided in the study, and to the instructors in the Department of Psychology at The Pennsylvania State College and the officers at the Great Lakes Training Station, Great Lakes, Illinois, for making test populations available.

A variety of secondary questions were studied in connection with this central one. Of these, the principal one was: Is there a diminution in interest as the length of film sessions is increased? If so, What relationship is there between learning and interest as film length is increased?

To obtain an answer to these questions, two independent experiments, one with eleven classes of undergraduate psychology students and the other with ten companies of Navy recruits, were undertaken.

PROCEDURES

Psychology classes' experiment.—The first experiment was the one employing the psychology students. The eleven classes each included between twenty-nine and sixty students, making a total sample in excess of four hundred. For this population, two series of films were selected. One was W. N. Kellogg's *Ape and Child* Series. The first four titles in this series were used. These films present a comparative study of maturation and learning in a human infant and a chimpanzee.[2]

The other series was J. H. Masserman's *The Dynamics of an Experimental Neurosis.*[3] The four reels of the series constitute a unit on the dynamics of experimental neurosis in cats and therapeutic methods used in alleviating the neurosis.

The *Ape and Child* Series was shown during March, 1948;

---

[2] The titles used included: "Some Behavior Characteristics of a Human and a Chimpanzee Infant in the Same Environment," "Comparative Tests on a Human and a Chimpanzee Infant of Approximately the Same Age," "Experiments Upon a Human and a Chimpanzee Infant after Six Months in the Same Environment" and "Some General Reactions of a Human and a Chimpanzee Infant After Six Months in the Same Environment." These were all silent, black-and-white films. The average running time was 14.5 minutes.

[3] The titles used were: "Conditioned Feeding Behavior and Induction of Experimental Neurosis in Cats," "Effects of Environmental Frustrations and Intensification of Conflict in Neurotic Cats," "Experimental Diminution of Neurotic Behavior in Cats," "Active Participation in Establishing More Satisfactory Adjustment." These were all silent, black-and-white films. The average running time was fifteen minutes.

the *Cat Neurosis* Series during April, 1948. Three classes were each shown each series in a single long session, lasting approximately one hour (1-Part Method). Two classes were shown each series in two sessions, two reels per session (2-Part Method). The sessions lasted approximately thirty minutes each, on alternate days. Two classes were each shown each series in four sessions, one reel per session (4-Part Method). The sessions lasted approximately fifteen minutes each, on alternate days. All sessions were held during regular class periods. With one exception no class was shown both series in the same spacing pattern.

Immediately at the end of a showing, of one, two, or four reels, the group was asked to fill out a rating form intended to yield some information about the effect of film length on interest.

One week from the midpoint of the showings, in the case of the *Ape and Child* Series, a 78-item multiple-choice test based on the film series was administered to the seven experimental classes, and to four classes who served as a control group. The latter took the test without having seen the films.

In the case of the *Cat Neurosis* Series, an 80-item multiple-choice test based on the film series was given to one class in each methods group two weeks after the mid-point of the showings, and to the other classes one week after the midpoint of the showings. The four classes who served as a control group for the *Ape and Child* Series took this test also.[4]

The comparability of the eleven classes with respect to initial status was determined using as criteria the all-college grade point average and the final psychology course grade of each subject.

The Navy experiment.—The second experiment was conducted in August, 1948, at the Great Lakes Naval Training Station in Illinois. In all, ten companies of recruits, each numbering between sixty and one hundred men, were used.

Two film series were selected, each comprising three reels. One was a series on *Rules of the Nautical Road.* The rules covered dealt with "Visual Day Signals," "Whistle Signals for Approaching Steam Vessels," and "Lights of Vessels Being Towed." The other series was on *Elementary Hydraulics*. The reels used included "Applications of Pascal's Law," "Characteristics of Liquids in Motion" and "Simple Hydraulic Systems." These were all sound films, using animation throughout. The average running time was fifteen minutes.

---

[4] The reliabilities of the tests were estimated by the Kuder-Richardson method of rational equivalence. Use was made of their Formula 20. The reliability of the *Ape and Child* Series test, based on the scores for all subjects in the experimental group, was .51. The reliability of the *Cat Neurosis* Series test was .73.

For each series 45-item multiple-choice tests were constructed.[5]

Five of the ten companies were assigned to each film series. For each series, two companies saw the three reels in a single session lasting about forty-five minutes (1-Part Method), two companies saw the three reels, one reel at a time, in three sessions, each session lasting about fifteen minutes (3-Part Method). One company served as a control group, taking the test without seeing the films. In the spaced or 3-Part Method groups, the intervals between the sessions could not be held constant. They were approximately twenty-four hours long, but for one company the two intervals were three hours and two days, respectively. No company participated in sessions for both film series.

As in the psychology replication, at the end of each session the group was asked to fill out a rating form, and one week after the midpoint of the series each company was tested.

The comparability of the companies with respect to initial status was determined using as criteria scores on the Navy General Classification Test and the Navy Mechanical Aptitude Test.

FINDINGS

Comparability of the groups.—Table 1 presents the means and standard deviations for the control and experimental groups for each film series, for the two matching variables used in each case.

The psychology groups were somewhat more variable with respect to the mean final psychology course grade than with respect to the grade-point average. In the former case, the differences among the groups were significant at the five per cent level of confidence, for both film series. The differences in grade-

---

[5] The Kuder-Richardson reliabilities for these tests were: *Rules of the Nautical Road* Series test, .57; *Elementary Hydraulics* Series test, .80.

TABLE 1.—MEANS AND STANDARD DEVIATIONS FOR THE MATCHING VARIABLES FOR THE CONTROL AND EXPERIMENTAL METHODS GROUPS IN EACH FILM SERIES GROUP

A. Psychology Classes Population

Method	Ape and Child Film Series					Cat Neurosis Film Series				
	No. of Cases	Grade-Point Avg.		Final Psychology Grade		No. of Cases	Grade-Point Avg.		Final Psychology Grade	
		Mean	SD	Mean	SD		Mean	SD	Mean	SD
Control	105	1.40	.04	1.21	.96	159	1.39	.04	1.23	.88
1-Part	111	1.36	.06	1.41	.98	108	1.44	.69	1.27	.93
2-Part	101	1.25	.64	1.44	.84	69	1.45	.56	1.52	.93
4-Part	83	1.50	.05	1.55	.91	73	1.33	.54	1.52	.81
		F = 2.44		F = 3.01*			F < 1		F = 2.90*	

B. Navy Population

Method	Elementary Hydraulics Film Series					Rules of the Nautical Road Film Series				
	No. of Cases	Gen. Class. Test		Mechanical Aptitude Test		No. of Cases	Gen. Class. Test		Mechanical Aptitude Test	
		Mean	SD	Mean	SD		Mean	SD	Mean	SD
Control	109	52.79	8.76	49.51	9.23	84	50.50	11.03	48.92	10.67
1-Part	190	51.38	10.77	48.78	10.21	193	51.49	10.53	49.75	10.39
3-Part	152	50.78	9.42	48.97	9.32	159	49.55	10.14	48.20	8.78
		F = 1.34		F < 1			F = 1.49		F = 1.06	

* Significant at the five per cent level of confidence.

point average were not significant at the five per cent level. Furthermore, the classes within each group (experimental method or control) did not depart significantly from homogeneity with respect to either the grade-point average or the final psychology grade.

There was no evidence that the Navy groups (or the two companies within each group) differed significantly either in mean General Classification Test score or mean Mechanical Aptitude Test score.

Inter-methods differences.—For all four film series, the inter-methods differences in total test score, at the end of one or two weeks, may be attributed to chance fluctuation. Table 2 presents the film test means, standard errors, and standard deviations for the experimental and control groups for each film series. In each case, the F-ratio for the experimental methods groups falls short of significance at the five per cent level of confidence.

The difference between the experimental and control groups, however, is in each case significant at the one per cent level of confidence.[6]

Finally, for the *Cat Neurosis* Series a comparison was made between the one-week and two-week retention groups. This difference was significant at the one per cent level of confidence (F-ratio of 12.78). It may be noted that the two-week retention group was somewhat more variable than the one-week group, and that the 1-Part Method group, when tested at the end of two weeks, showed a loss relatively greater than that for either of the other two methods groups.

An analysis was undertaken of the results on the subtests (one for each reel in each series). In general, the conclusions based on the total test score apply equally to the subtest scores. A highly significant difference was found between the experimental and control groups, and in almost all cases no significant difference among the methods was noted. However, for one subtest for the *Cat Neurosis* Series and for two subtests for the *Ape and Child* Series, the analysis indicated variation greater than can be accounted for by chance factors alone. In one of the *Ape and Child* Series subtests the difference in means was in favor of the massed (1-Part) presentation; in the other two instances the difference was in favor of the spaced presentations. Since these findings are based on subtests including only twenty items, and since the bulk of the evidence is in the other direction, they cannot be considered as seriously disturbing the stability of the principal conclusion.

---

[6] These F-ratios were as follows: *Ape and Child* Series—2428.010; *Cat Neurosis* Series—502.202; *Rules of the Nautical Road* Series—30.416; *Elementary Hydraulics* Series—63.836.

TABLE 2.—SUMMARY OF FILM TEST SCORES: MEANS, STANDARD ERRORS OF MEANS, AND STANDARD DEVIATIONS

A. Psychology Classes Population

Method	Ape and Child Series				Cat Neurosis Series			
	No. of Cases	Mean	SE_m	SD	No. of Cases	Mean	SE_m	SD
Control	165	31.6	.41	5.2	159	22.5	.33	4.2
One-Week Retention Group								
1-Part	111	54.3	.44	4.7	62	38.6	1.09	8.5
2-Part	101	55.1	.44	4.4	37	39.3	1.31	7.8
4-Part	83	55.0	.52	4.7	23	40.7	1.46	6.9
		$F < 1$†				$F < 1$†		
Two-Week Retention Group								
1-Part					47	34.2	.69	8.2
2-Part					32	37.1	1.49	8.3
4-Part					50	37.3	1.21	4.8
						$F = 2.63$†		

B. Navy Population

Method	Rules of the Nautical Road Series				Elementary Hydraulics Series			
	No. of Cases	Mean	SE_m	SD	No. of Cases	Mean	SE_m	SD
Control	84	12.9	.33	3.0	109	14.7	.23	4.8
1-Part	193	16.5	.35	4.8	190	19.5	.22	7.3
3-Part	159	15.7	.34	4.2	152	20.1	.25	6.2
		$F = 2.50$†				$F = 1.05$†		

† F-ratios refer to the experimental methods groups only. The comparisons between the experimental and control groups are given in the text.

The effect of initial status.—It has already been demonstrated that the groups did not depart significantly from homogeneity, except perhaps in one instance, with respect to the 'initial status' or matching variables, considered individually. The question remained—When these variables (grade-point average and final psychology grade for the psychology groups, Navy General Classification Test and Mechanical Aptitude Test for the Navy groups) are used jointly to define initial status, and the groups are matched by means of an analysis of covariance, do significant differences emerge as between the presentation methods?

It was found, first, that the multiple correlation between the film test score and the two matching variables was generally low.[7] Second, when the variance estimates were adjusted to take into account the regression with the matching variables, there was for each test a slight reduction in the estimate of error, but in all cases the F-ratios remained well below the five per cent level of confidence. In short, the lack of differences among the experimental methods cannot be considered an artifact of initial differences among the groups, at least with respect to these matching variables.

The interest ratings.—The interest rating questionnaires were devised and administered on the hypothesis that, even though the massed film presentation might not yield less learning than spaced presentations, subjects in the massed (1-Part) groups would report less interest than subjects in the spaced groups. It was thought that the satiety achieved in one long period would result in a higher frequency of negative responses to such questions as, "Would you like to see more films in this series?" and "Did these films hold your interest all the time?", than would be the case in the spaced presentations.

This hypothesis was not consistently sustained by the findings. Table 3 presents the mean ratings for each session for each experimental method for the four-film series. In the Navy population,

---

[7] These multiple correlations were as follows: between *Ape and Child* Series total score, and grade-point average and final psychology grade .20; between *Cat Neurosis* Series total score, and grade-point average and final psychology grade .23; between *Rules of the Nautical Road* Series total score, and Navy General Classification Test and Mechanical Aptitude Test .42; between *Elementary Hydraulics* Series total score, and Navy General Classification Test and Mechanical Aptitude Test .02.

TABLE 3.—MEAN RATINGS AND CORRELATIONS BETWEEN RATINGS AND FILM TEST SCORES

A. Psychology Classes Population

Method	Film Session	Ape and Child Series			Cat Neurosis Series		
		Mean Rating	SD	r*	Mean Rating	SD	r*
1-Part	First (only)	15.1	1.9	.00	11.1	2.5	.20
2-Part	First	15.1	1.6	.09	13.1	2.8	.43
	Second	15.2	1.6	−.10	12.7	2.9	.25
4-Part	First	14.4	1.6	−.07	13.6	1.7	.02
	Second	14.2	1.4	−.07	12.1	2.0	.08
	Third	14.1	1.4	.19	13.3	2.1	.13
	Fourth	14.4	1.4	−.01	13.4	1.9	.10

B. Navy Population

Method	Film Session	Rules of the Nautical Road Series			Elementary Hydraulics Series		
		Mean Rating	SD	r*	Mean Rating	SD	r*
1-Part	*First (only)*	3.5	1.4	.16	3.3	1.6	.09
3-Part	First	3.5	1.4	.14	3.4	1.5	.17
	Second	3.6	1.4	.03	3.9	1.4	.16
	Third	3.9	1.4	.11	3.9	1.4	.11

* Correlation between rating and score on the subtest (or tests) for the film or films shown during the period in which the rating was made.

the mean rating scores for the massed session were slightly lower than the means for the spaced sessions, by a fraction of a point. However, for the *Rules of the Nautical Road* Series, the differences among the massed session and the first and second spaced sessions were not significant, while the third spaced session was given a significantly higher mean rating (3.9) than either of the other two spaced sessions or the massed session (3.5, 3.6, and 3.5, respectively). In the case of the *Elementary Hydraulics* Series the mean ratings for both the second and third spaced sessions (3.9 in both cases) were significantly higher than the means for the massed session (3.3) or the first spaced session (3.4).

In the psychology classes population, for *the Ape and Child* Series the means for the massed and the 2-part spaced sessions were both higher than the means for the 4-part spaced sessions, while the reverse was true for the *Cat Neurosis* Series. Furthermore, analysis of the data showed that intra-session variability (differences between the classes or Navy companies exposed to the same treatment) was generally as great as, or greater than, inter-session or inter-method variability.

When the scores on the film tests were correlated with the ratings, a near-zero relationship was found. The distribution of responses to the specific questions on the rating form also failed, except in one instance, to provide any consistent differentiation among the methods. This exception was provided by the data on the *Cat Neurosis* Series. Of the subjects seeing this series, sixty to eighty per cent in each class in the massed (1-part) group rated the session 'too long,' and thirteen to twenty per cent rated the two-part sessions 'too long.' For the other series, no more than ten per cent, and usually five per cent or less, rated any session 'too long.'

In general, then, the ratings did not afford a means of differentiating among the experimental groups, and interest, as rated, was relatively independent of learning accomplished.

SUMMARY AND CONCLUSIONS

Two four-reel series of instructional films were shown to undergraduate psychology students. Each series was presented in three different spacing patterns: four reels in one hour, two reels in each of two half-hour periods, and one reel in each of four fifteen-minute periods. Two three-reel series of films were shown to Navy recruits, in two spacing patterns: three reels in one forty-five-minute period, and one reel in each of three fifteen-minute periods. All experimental groups were tested, one or two weeks after the films were shown, and each group rated the films at the end of each session. In addition, for each series a control group that had not seen the films was given the test.

Inter-methods differences.—For all four series, the inter-methods differences in total test scores, at the end of one or two weeks, may be attributed to chance fluctuation. However, the difference between the one-week retention group and the two-week retention group was significant at the one per cent level of confidence. In general, these results apply equally to the scores on the subtests. For every series, the difference between the control and experimental groups was highly significant, indicating that learning took place. The lack of differences among the presentation methods could not be attributed to initial differences in the groups.

Interest ratings.—In general, intra-session variation and inter-company variation were as great as inter-method variation. Analysis of the distribution of responses to the questions on the interest rating form failed to provide any consistent means of differentiating among the methods. However, for the *Cat Neurosis* series between sixty and eighty per cent of the subjects rated the one-hour (massed) session 'too long.' Finally, the correlations between ratings and test scores were about zero. The data suggest that the ratings reflect other factors than the adequacy of the films as teaching devices.

IMPLICATIONS

First, the evidence suggests that training sessions devoted to films may last as long as an hour with the expectation that sustained learning comparable to that achieved when several short film sessions are employed will occur. In mass training programs considerable economies in scheduling may be realized on this basis.

Second, for essentially homogeneous content producers may profitably explore the relative advantages of making a few long films rather than a large number of single reels.

Finally, these results pose an interesting question with respect to the limits of massing material. It is not unusual for college class lectures to last two hours or more. What would the consequences be if films were lengthened to that extent, or more? To what extent is this apparent lack of difference affected by varying subject-matter difficulty, the intervals between film sessions, and other factors?

A STUDY IN PROGNOSIS: THE GUIDANCE VALUE OF SELECTED MEASURES OF MUSICAL APTITUDE, INTELLIGENCE, PERSISTENCE, AND ACHIEVEMENT IN TONETTE AND ADAPTION CLASSES FOR PROSPECTIVE INSTRUMENTAL STUDENTS

HAROLD CARL MANOR
Department of Music, Arkansas State College

INTRODUCTION

Research in the field of musical talent and its relation to achievement in music study has revealed that the factors which contribute to or determine success are at the same time both complex and manifold. Musical talent, because of its complexity, probably cannot be measured by any one test or scale.

This study is primarily concerned with the initial guidance of prospective instrumental students, and it proposes to suggest one solution to the problem by improving the selection of such pupils through the use of a broad profile of physiological, mental, and other psychometric measurements.

One of the objectives of the study is to open the area of adaption or clinic classes as a means of selecting instrumental pupils for further study. A second objective is to further study the value of the tonette and a measure of persistence in such selection of pupils.

The problem was to ascertain the prognostic value of selected measures of musical aptitude, intelligence, persistence, and achievement in tonette class and adaption classes as revealed by a special technique, in guiding fourth-grade pupils into instrumental classes.

SUMMARY OF THE LITERATURE

The literature which deals directly with the several relationships encompassed by this study may be summarized as follows:

1) The median of six correlations reported between the Seashore tests and success in instrumental and vocal performance was .375.

2) The Tonal Memory Test of the Seashore tests correlated .24 with success on woodwind instruments, .28 with success on string instruments, and .45 with success on brass instruments.

3) The median of eight correlations reported between various measures of intelligence and musical achievement was .38. These correlations ranged from .20 to .50.

4) Three correlations with a median of .24 were reported between the Seashore tests and measures of intelligence.

5) The median of five coefficients of contingency, ranging from .37 to .47, between persistence and musical achievement was .39. One correlation of .68 between persistence and musical achievement was reported.

6) Physical measurements of finger length for players of the cornet, clarinet, and violin correlated with success on these instruments from −.13 to .38 with a median of .15. Correlations between success on the cornet and clarinet and measures of teeth evenness were −.01 and .13, respectively.

7) The adaption class and clinic class is recommended by writers in the field of music education.

8) Correlations were reported for one battery of tests, including the Seashore tests and measures of intelligence. The correlations were .58 between the battery and success on the cornet, .42 between the battery and success on the clarinet, and .39 between the battery and success on the violin.

MATERIALS AND PROCEDURES

The special technique of the study was the organization of the program of testing and instruction as an exploratory course. Selected fourth-grade classes were given the Pitch, Tonal Memory, and Rhythm tests of the Revised Seashore Tests of Musical Talent, and the California Elementary Intelligence test, Short Form, for grades IV to VIII, and each pupil was ranked on the Manor Persistence Ranking Scale. During the eight-week period devoted partly to testing, the pupils in these fourth-grade classes received instruction on the tonette. At the end of a period of eight half-hour periods of instruction on the tonette (one a week) the pupils were given a sight reading test. All pupils in each room then received instruction on the violin, clarinet, cornet, trombone, and drum. Score cards were designed in an attempt to objectify the teacher's judgment as to which instrument the pupil might begin to study with an initial advantage.

After this adaption class, those pupils who were able to purchase an instrument, or for whom the school could furnish one, entered the instrumental class. The instrumental class covered a period of fourteen weeks, or fourteen lessons at the rate of one lesson per week. A final achievement test consisting of a repetition of the tonette test played on the instrument studied, a second score on the adaption score card for the instrument studied, a sight reading test on the practice test, and, after practice for one week, another score on the practice test were given to each instrumental pupil.

The reliability of the persistence ranking was established by correlating two rankings of a class of pupils by each of four teachers. The re-rank coefficients ranged from .72 to 1.00, with a median of .94. The validity of the persistence ranking was determined by correlation with an accomplishment quotient. The correlations for four accomplishment quotients obtained in four subject-matter areas ranged from .32 to .59, with a median of .45.

The reliability of the tonette test was determined by the split-half method. The split-half reliability coefficient was found to be .88, and, by using the Spearman-Brown formula, the reliability of the whole test was found to be .94.

The adaption class score cards were designed and submitted to authorities in the field of instrumental music education for approval. Final revisions were made and used on a demonstration class by five graduate students. The agreement in score rank was close, seldom varying over two or three points, except for the violin scores which had been given first. In the opinion of the five graduate students, close agreement in violin scores would have resulted if the violin score card had followed one of the other tests.

The adaption score cards provided opportunity to award points for tone production, range, fingering, physical execution, tone quality, and interest in that instrument. The number of successful trials and performance that matched described levels were used to obtain differences in score points.

The practice or sight reading test of the final achievement tests was given to thirteen elementary-school, nineteen junior high school, eighteen senior high school, and ten college instrumental pupils as a sight reading test. The average differentiation in score for the violin, clarinet, cornet, trombone, and drum tests between the four levels of instrumental pupils was six score points or an additional six measures played correctly by each higher performance ability group of instrumental pupils.

The musical material included in the final test was selected from the instruction books used in the instrumental class and presented increasingly difficult performance and musical problems.

All tests were either mimeographed in black on white or blue-printed. Copies of each test, with instructions for administering and scoring with class record blanks, were included in a manual of instructions for each teacher. All tests were sent to the director of the study for rescoring, and a list of the scores was returned to the teacher.

The original sample included twenty-four classes of fourth-grade pupils from Evansville, Jasper, Huntington, and Bloomington, Indiana, and Champaign, Illinois. Eight classes, consisting of two hundred seventy-three pupils, were unable to complete the exploratory program and were discarded or dropped from the study. Of the sixteen remaining classes, seven, consisting of one hundred eighty-four pupils, were not able to complete the instrumental class or did not have any pupils in it. Nine, consisting of two hundred nineteen pupils, contributed fifty-eight instrumental pupils to the study.

FINDINGS AND CONCLUSIONS

Organization of the Program.—Comment upon the organization of the program by the teachers participating in the experiment was a part of the data to be obtained from the study. The comment as summarized below should be of value to future attempts to organize such a program.

1) All of the ten teachers participating in the program expressed a desire or need for an organized program of instrumental guidance. Most of these teachers had been searching for or designing such a program for their own schools.

2) Only the two teachers who taught in one school were able to complete the entire program during the year as scheduled in the manual of instructions. The teachers in the Evansville schools,

with one exception, taught in two or more schools but used only one class each in the experiment.

3) The program or schedule of the teachers assigned to more than one school was so inflexible that if periods were missed they could not be made up. Individual pupils missing tests were not able to meet with the teacher at special times to take the tests missed.

4) Several of the schools had scheduled the program for Friday and found that many Fridays were missed in their school. These teachers would recommend that a one-period-a-week class be scheduled sometime during the middle of the week.

5) All of the teachers who were able to present at least one of the adaption classes commented upon the fact that interest in participation in instrumental music was greater among the pupils after these pupils had taken part in the adaption class.

6) The three teachers completing the adaption classes for their schools seemed to feel that about eighty per cent of the pupils entering the instrumental class had chosen the instrument that the teacher would have selected for that pupil. The balance of instruments, except for the violin, seemed to be satisfactory for each school with planned changes from cornet to French horn at a later date.

7) Some schools plan to use the program, somewhat modified, as a year program in the fourth grade, starting all instrumental pupils in the fifth grade.

8) Any exploratory program, such as the one used in this study, apparently should be organized in terms of the needs and facilities of the school using it.

Relationships for the Whole Group. The relationships found between the various guidance measures for the whole group are probably a closer approximation to the true relationships between these factors than the relationships found with only the instrumental group because of a much smaller number of cases and the possible bias in the instrumental group scores due to factors of selection operating. The following points are indicated by a study of the data for the whole group:

1) The discarded classes varied significantly on the basis of an analysis of variance of the non-language intelligence scores. A value for F of 602.3 was found for the variance in these scores. A value for F of 5.70 would be exceeded by only one per cent of all

similar random samples. Consequently the observed value of 602.3 indicates a highly significant variance.

2) A comparison of the discarded and retained groups of pupils on their previous training in instrumental music reveals that thirty-six per cent of the retained group and only five per cent of the discarded group had previous training.

3) The retained and discarded groups were compared on the basis of their respective score distributions on the tonette, violin, Seashore Pitch, and total intelligence tests. The chi-square test was used and significant values for chi-square above the one per cent level of probability of occurrence by chance were found for all but the tonette scores. The value of chi-square found for the difference between the two tonette score distributions would be exceeded by not less than one per cent and not more than five per cent of all random samples.

TABLE I.—THE CHI-SQUARE TEST APPLIED TO THE TOTAL INTELLIGENCE SCORES OF THE DISCARDED AND RETAINED GROUPS

Interval Code	Retained Group	Discarded Group	Per cent t	t	$o - t$	$(o - t)^2$	$\frac{(o - t)^2}{t}$
8	1	1	.66				
7	24	3	22.44				
6	57	16	37.62	60.72	−40.7	1656.49	27.29
5	110	55	72.60	72.60	−17.6	309.76	4.26
4	49	69	32.34	63.36	27.6	761.76	12.01
3	96	91	63.36	32.34	36.7	1346.89	41.70
2	17	13	11.22	11.22	1.8	3.24	.30
1	1	8	.66	25.08	−16.1	259.21	10.32
0	37	1	24.42				
Total	392	257					$X^2 = 95.88$

Table I shows the chi-square test applied to the total intelligence scores of the discarded and retained groups. From a table for P for five degrees of freedom a value of 15.086 would be exceeded by only one per cent of all random samples. The obtained value of 95.88 in Table I is highly significant and the difference between the total intelligence scores of the discarded and retained groups are larger than could be attributed to chance. In view of the significant differences observed between the discarded and retained groups, we must assume that different results might have been observed for the study if it had been possible to retain these pupils.

4) Because of the similarity between the tonette class and the instrumental class in the manner in which both are taught and measured, the tonette scores of the retained classes were used to analyze the variance between the sixteen rooms comprising the retained group. It was found that the differences between ten classes taught by the same teacher were significant although the differences between two classes in the same school from that group were not significant. The differences between six classes taught by different teachers were also significant. The differences between the sixteen classes of the retained group were also significant. However, when the differences between the ten classes taught by one teacher and the six classes taught by different teachers were analyzed, it was found that the differences were not significant.

TABLE II.—THE VARIANCE BETWEEN TEN CLASSES TAUGHT BY ONE TEACHER AND SIX CLASSES TAUGHT BY SIX DIFFERENT TEACHERS

	d.f.	S.S.	Variance
Between six classes	5	1807.41	361.48
Between ten classes	9	1984.05	220.45
Total	14	3791.46	F = 1.19

Table II shows the variance between ten classes taught by one teacher and six taught by six different teachers. From a table of F for nine and five degrees of freedom, we find an F of 4.78 would be exceeded in five per cent of all samples of this size and an F of 10.15 would be exceeded in one per cent of all samples of this size. The obtained F of 1.19 indicates that the teacher variable between these two groups of classes is insignificant.

It is reasonable to assume that since the other classes have been taught in a similar manner, the teacher variable for the rest of the experiment might have been of no particular significance.

5) An examination of the correlations in Table III reveals that the three measures of intelligence show about the same relation to the other measures in the study. The correlations between non-language intelligence and the other ten factors ranged from −.01 to .33 with a median of .13. The correlations between language intelligence and the other factors ranged from −.04 to .40 with a median of .10. The correlations between total intelligence and the other factors ranged from .00 to .43 with a median of .10. The correlations between non-language, language, and total intelligence and persistence were .33, .40, and .43, respectively, indicating the possibility that intelligent, as well as persistent, behavior has influenced the persistence ranking.

6) The Seashore Pitch, Tonal Memory, and Rhythm tests correlated .33, .41, and .33, respectively, with the violin adaption class score. The correlations between the three Seashore tests and the other adaption scores ranged from −.12 to .18 with a median of .07. The Rhythm test showed correlations of .36, .62, .62, and .40 with the first four subtests of the drum adaption class test. The Rhythm test scores correlated .52 with the tonette test scores, which indicated a considerably closer relation than either the Pitch or Tonal Memory tests showed. The correlations between the Pitch and Tonal Memory test scores and the tonette test scores were .17 and .15, respectively.

7) The intercorrelations between the three Seashore tests might indicate some overlap in the capacities measured by each test. The correlation between the Pitch and Tonal Memory tests was .60, and between the Pitch and Rhythm tests it was .37. The correlation between the Tonal Memory test and the Rhythm test was .37.

8) The best factor for predicting tonette success was the persistence ranking, with a correlation of .60. The tonette scores appear to be more closely related to the violin and drum adaption scores, as indicated by correlations of .39 and .27, respectively, than to the clarinet, cornet, and trombone scores as shown by correlations of .15, .05, and .05, respectively.

9) The intercorrelations between the adaption class scores range from .01 between the violin and clarinet scores to .61 between the cornet and trombone scores. The median intercorrelation of ten coefficients was .24.

10) The obtained correlation of −.77 between success on the

TABLE III.—INTERCORRELATIONS OF THE PREDICTIVE FACTORS

Measures	Pitch	Ton. Mem.	Rhythm	Persist.	Tonette	Violin	Clarinet	Cornet	Trombone	Drum
C.A.*	−.10	−.16	−.03	−.18	−.13	−.13	−.01	.07	.01	−.04
n	300	300	300	342	342	185	247	266	266	85
Non-L.	.15	.22	.11	.33	.28	.18	.05	.08	.11	−.01
n	300	300	300	342	342	185	247	266	266	85
Lang.	.18	.23	.00	.40	.37	.12	−.04	−.01	.00	−.01
n	300	300	300	342	342	185	247	266	266	85
Tot. IQ	.21	.27	.11	.43	.38	.25	.00	.02	.05	.00
n	300	300	300	342	342	185	247	266	266	85
Pitch		.60	.37	.11	.17	.33	.09	−.04	.14	.14
n		329	329	329	329	141	253	294	294	85
Ton. Mem.			.37	.12	.15	.41	.06	.00	.15	.08
n			329	329	329	141	253	294	294	85
Rhythm				.69	.52	.33	.00	−.12	.14	.18
n				329	329	141	253	294	294	85
Persist.					.00	.22	.10	.02	.08	.22
n					185	185	255	299	299	85
Tonette						.39	.15	.05	.05	.27
n						185	255	299	299	85
Violin							.01	.02	.23	.46
n							112	112	112	85
Clarinet								.20	.25	.35
n								258	258	85
Cornet									.61	.16
n									315	85
Trombone										.40
n										85

* CA is chronological age; Non-L. is non-language IQ; Lang. is language IQ; and Tot. IQ is the combined IQ; Ton. Mem. is the Tonal Memory Test of the Seashore tests; and Persist. is the Persistence ranking.

tonette and previous experience in months is not consistent with the experience of the teachers participating in the study. Experience in months does not indicate the value of the experience in terms of skills, musical knowledge, and ability, and it is the opinion of these teachers that the information rather than the relationship is at fault.

11) On the basis of the experience of the teachers participating in the study we must conclude that the same program cannot be administered equally well in every school. Any program designed for guidance purposes in instrumental music will likely have to be modified in terms of the school which it must serve. All of the teachers participating in the study have expressed the need for such a program and indicated that the adaption classes, where used, stimulated interest in instrumental music.

Correlations between Certain Factors Reported.—The relationships found for the instrumental group are in close accord with the findings of previous investigators, with the exception of the relationship observed between intelligence and instrumental achievement. The median of eight correlations reported between intelligence and musical achievement was .38, the median found in this study was .09. The median of six correlations between the Seashore tests and instrumental and voice achievement was .38, and in this study the median found was .32. The median of three correlations reported between the Seashore Tonal Memory test and instrumental achievement was .24, and for this study the median was .32. The median of six correlations reported between persistence and musical achievement was .40, and for this study the median was .36 with instrumental achievement and .44 with tonette achievement. The relationship found between the Seashore tests and intelligence in this study was .23, a median of three correlations, and the reported relation in the literature was .24, a median of three correlations.

Apparently there is a close agreement between the findings of this study and previously reported studies for these relationships.

Concerning the instrumental group. The following conclusions were reached concerning the relationships found between the achievement of the fifty-eight instrumental pupils and the guidance measures.

1) The correlations found between the non-language, language, and total intelligence measures and achievement of .17,

TABLE IV.—INTERCORRELATIONS BETWEEN THE VARIOUS GUIDANCE MEASURES AND CORRELATIONS BETWEEN THE GUIDANCE MEASURES AND THE ACHIEVEMENT TESTS

Measures	1 Non-Lang.	2 Lang.	3 Tot. IQ	4 Pitch	5 Ton. Mem.	6 Rhythm	7 Persist.	8 Tonette No. 1	9 Adaption No. 1	10 Practice Test	11 Sight Reading Total	12 Previous Training	13 Practice Time	14 Achievement Score
1		.40	.81	−.02	.29	.18	.13	.14	.04	.10	.07	.12	.35	.17
2			.79	−.01	.23	.26	.37	.20	−.15	−.02	.04	.33	.05	.05
3				−.01	.38	.26	.26	.12	−.12	.01	.05	.26	.22	.09
4					.45	.37	.22	.26	.19	.45	.29	.30	−.11	.49
5						.28	.16	.00	.16	.29	.21	.24	.16	.32
6							.27	.06	−.03	.21	.11	.17	−.06	.21
7								.44	.12	.27	.17	.22	−.20	.36
8									.13	.33	.24	.18	.17	.41
9										.35	.14	−.01	−.06	.27
10											.49	.20	.22	.88
11												.14	.09	.56
12													−.12	.29
13														.13

.05, and .09, respectively, would have justified dropping these measures of intelligence but leaving the whole question of intelligence test measures open for further investigation.

2) Of the three Seashore tests, the Pitch Test showed the closer relation to the achievement scores, a correlation of .49, also the highest correlation found between the guidance measures and achievement. The Tonal Memory test of the older Seashore tests usually produced the higher correlation between the Seashore tests and measures of achievement. The correlation found between the achievement scores and the Tonal Memory test in this study was .32, and a correlation of .21 was found for the Rhythm test, as shown in Table IV.

3) The persistence ranking was significantly related to pupil success on both the tonette and an instrument. The correlation between the persistence ranking and instrumental achievement scores was .36 and between persistence and tonette success it was .44.

4) Success on the tonette produced the second highest correlation with instrumental achievement, an r of .41. This relationship would indicate that success on the tonette is as valid a guidance measure as the Seashore tests for guiding prospective instrumental pupils.

5) *Previous training and the adaption class scores* were of about equal value in predicting instrumental success as is indicated by their respective correlations with achievement of .29 and .27. Previous elimination and self-guidance, based somewhat upon the adaption class scores, may have placed too great a burden upon the adaption measures in expecting them to distinguish between success levels. It might well be that with a refinement of the measurements in the adaption class scores, such discrimination would be possible.

As the quartile disposition of achievement scores given in Table V would seem to indicate, the adaption class scores were about as effective in selecting the highest and lowest quartiles of achievers as the other guidance measures. It was found that fifty per cent of the first quartile of the achievement scores were drawn from the first quartile of the adaption class scores and that forty-three per cent of the fourth quartile of the achievement scores were drawn from the fourth quartile of the adaption class scores. The Seashore Pitch test showed that fifty-seven per cent of the first

TABLE V.—THE PER CENT OF EACH ACHIEVEMENT QUARTILE DRAWN FROM EACH QUARTILE OF THE TEN PREDICTIVE MEASURES

Predictive Measure Quartile	Achievement Quartile	Non-language IQ	Language IQ	Total IQ	Pitch	Tonal Memory	Rhythm	Persistence	Tonette No. 1	Adaption No. 1	Previous Training on Instrument
q1	Q1	.22	.21	.21	.57	.50	.37	.36	.50	.50	.36
	Q2	.13	.27	.20	.53	.33	.40	.27	.27	.20	.20
	Q3	.40	.20	.27	.07	.00	.07	.07	.13	.13	.13
	Q4	.21	.29	.29	.14	.21	.07	.21	.14	.14	.07
q2	Q1	.43	.21	.29	.22	.14	.14	.36	.36	.22	.21
	Q2	.34	.33	.34	.07	.27	.27	.33	.33	.40	.27
	Q3	.13	.20	.27	.27	.47	.45	.13	.20	.27	.20
	Q4	.21	.36	.21	.14	.07	.21	.14	.14	.14	.29
q3	Q1	.22	.29	.36	.14	.14	.28	.14	.07	.22	.43
	Q2	.13	.27	.12	.20	.20	.27	.13	.20	.33	.53
	Q3	.27	.27	.27	.39	.40	.21	.40	.20	.13	.67
	Q4	.37	.14	.36	.29	.29	.36	6	.58	.29	.64
q4	Q1	.14	.29	.14	.07	.22	.21	.07	.07	.07	*
	Q2	.40	.13	.34	.20	.20	.13	.27	.20	.07	
	Q3	.20	.33	.19	.27	.13	.27	.40	.47	.47	
	Q4	.21	.21	.14	.43	.43	.36	.29	.14	.43	

* Thirty-three pupils or fifty-seven per cent of the entire group indicated no previous training. The twenty-five pupils indicating previous experience were divided into two groups, one of eleven and one of fourteen. These were labeled quartile 1 and 2, respectively. The group of thirty-three indicating no experience were grouped into one quartile space for this table.

quartile and forty-three per cent of the fourth quartile of achievement scores were drawn from the first and fourth quartiles, respectively, of the Pitch scores. The Tonal Memory test of the Seashore tests showed the same per cent from the first and fourth quartiles as the adaption scores. The other guidance measures showed smaller percentages at one or both extremes than the adaption class scores.

6) The smaller number of cases in the quartile distributions for the three teachers would make a direct comparison of the effectiveness of the measures for each teacher inconclusive. The quartiles contained from four to six pupils' scores and the change of one score would change the per cent from seventeen to twenty-five points in each instance. The per cent found for the three teachers was averaged and compared with the per cent in each quartile found for the group as a whole. The difference was usually within from five to ten per cent, indicating the possibility of some consistency in the operation of these measures in the experiment.

7) Although the measures might be helpful in a guidance program, rigid differential individual guidance cannot be justified on the basis of the results found in this study.

RECOMMENDATIONS FOR THE USE OF THE STUDY

None of the obtained correlations were sufficiently high that rigid differential individual guidance could be justified. However, the following points seem indicated from a review of the experiment.

1) The correlations between the non-language, language, and total intelligence scores and achievement were .17, .05, and .09, respectively. These results would seem to indicate that the intelligence measures might well be omitted from the testing program.

2) The correlations found between the Pitch, Tonal Memory, and Rhythm tests of the Seashore battery which were .40, .32, and .21, respectively, would indicate that, although individual guidance on the basis of these measures would not be justified, they may be helpful in pointing out pupils for consideration who might otherwise be overlooked.

3) The correlation of .36 found between persistence and achievement and of .41 between the tonette scores and achievement would indicate that both of these measures might be of

some aid to the teacher. The pupil-teacher association in the tonette class gives the teacher an opportunity to know better the pupils with whom he is working in addition to the guidance value of the tonette class.

4) The adaption score cards were not able to differentiate clearly between those pupils who were among the middle success scores. However, since preliminary differentiation had been made in many cases by the operation of the adaption class, the adaption measures may be more effective than the results show. It is reasonable to assume that the initial differentiation would be more important to both pupil and teacher than the final difference in success levels.

5) The three instrumental teachers would have had about a 50-50 chance of predicting the upper and lower quartiles of achievement scores from the upper and lower quartiles of the predictive measures. Slightly better than chance predictions of the upper and lower halves of the achievement scores could have been made. Considerable caution must be exercised in the use of these measures for guidance purposes, and possibly others substituted.

RECOMMENDED CHANGES OR MODIFICATIONS

Each of the changes proposed in the present study will be explained as they are presented.

1) The three intelligence measures might be dropped from the testing program. This would seem to be indicated by the correlations found in this study between these measures and achievement. The correlations ranged from .05 to .17.

2) The Seashore Rhythm test might be dropped from the testing program also. The correlations found between the Seashore Rhythm test and the subtests of the drum adaption class test, which range from .36 to .62, indicate considerable relationship. Since the drum adaption class serves at least one other function, that of introducing the pupil to the drum itself, and, the drum scores and the Rhythm test seem about equally related to achievement as indicated by correlations of .27 and .21, respectively, it is recommended that the drum adaption class test serve as the rhythm measure of the testing program.

3) It is recommended that the music teacher rank the pupils during the tonette class at least once, and preferably twice, in

addition to the room teacher's ranking. The average of these three rankings might prove a better estimate of the persistence of the pupil.

4) The tonette test given at the end of the tonette instruction was a sight reading test. This measure should be broadened to give a better measure of the musical knowledge gained. The test might include a written test on the knowledge of musical symbols used in the instructional material. Some credit might be given for the ability to play a memorized melody correctly although the pupil cannot read the notation for that melody. Some measure of the facility with which the tonette is played and the control exercised upon the tonette should be included in the test.

5) It is recommended that wherever possible, the length of time spent upon each instrument in the adaption classes be increased. The limit of diminishing returns cannot be set from this investigation, but from teacher comment it is recommended that the class be extended until the teacher is reasonably certain that each individual pupil has ceased making marked improvement over the previous day's efforts.

6) A performance test combined with a written test might prove to be a better measure of previous musical experience than merely the length of time a pupil has played an instrument.

7) The intercorrelations between the various parts of the achievement test would indicate that all of these parts might be retained. The correlations range from .13 to .49.

ADDITIONAL RESEARCH SUGGESTED

One of the objectives of this study was to open the area of the clinic or adaption class for further investigation. Some of the problems that seem to need further study are presented, with suggestions that might prove helpful in such investigations.

1) The tests that make up the score cards for the several instruments may not be the best group of tests for that instrument. There is a need for a larger sampling of tests for each instrument also. From a large sample of tests for one of the score cards, a better group of tests for guidance purposes might be selected. Using the violin score card as an example, an outline of an investigation is given here.

Each phase of violin technique covered in the violin score card could be expanded. The bowing skills might include six patterns instead of one pattern. This might be whole bow, half bow, whole bow followed by two bows in the upper half of the bow and the reverse, broken bows in the quarter and eighth note pattern of 6/8 time, plus additional string crossing exercises. The fingering exercises might be repeated on the A and G strings. Pictures might serve to demonstrate position as well as verbal description, and recordings might aid in establishing an aural picture of tone quality for comparison.

From repeated testings during instructional classes for the violin, an analysis of the improvement for the several test could be made which would indicate the tests which differentiate between the high and low total scores consistently and the weight in the total score to be assigned to each test. The minimum duration of the adaption class which would still provide differential measurement might be ascertained. This type of investigation might be made for each of the score cards.

The small number of pupils available for each type of instrument in this study prevented an analysis of the effectiveness of each score independently. If one class having over thirty-five pupils could be used, for example, in which all of the pupils would play one instrument, a closer approximation of the guidance value of the score card could be obtained.

2) Experimental data on the value of the tonette for purposes of instrumental guidance is practically non-existent. The relationship found between success on the tonette and achievement in instrumental music would seem to justify further investigation. The test for tonette achievement could include a broad measure of skills and ability that might point to the optimum period of instruction for guidance purposes. The number of zero scores found for many pupils in this study might indicate that the period of instruction was not long enough to obtain an adequate measure upon the pupils who apparently could not learn to sight read the material required in the test.

3) Further investigation of the Revised Seashore Tests of Musical Talent may reveal different relationships than those generally found for the older tests. The Pitch test seemed to show closer relation to achievement in this study, whereas the Tonal Memory test of the older measures usually correlated

more highly with measures of achievement than the rest of the tests.

4) The conflict in schedule between special classes that meet once or twice a week and special events such as convocations, vacations, teachers' meetings, etc., suggest that the problem might be investigated in order that these special classes might be scheduled at a time when fewer conflicts would exist.

5) The problem of evaluating musical experience on an instrument for guidance purposes may be closely related to the problem of the guidance value of tonette experience. Some measure of performance, of ability to read music notation, of attitudes, and musical perception probably should be included along with the rate of development of the pupil before evaluation is made. A series of case studies might disclose the technique that would enable evaluations of larger groups of pupils to be made adequately.

It is hoped that other problems in addition to the ones outlined above will suggest themselves. It is realized that the present study is only a beginning and that, before the value of the new measures used in this study and of other possible measures not employed by us can be ascertained, additional studies will have to be made. The dearth of research in the field of the tonette and adaption class must be remedied before certain forward progress in instrumental guidance can take place.

BIBLIOGRAPHY

1) Mary E. Allen, "A Comparative Study of Negro and White Children on Melodic and Harmonic Sensitivity," *Journal of Negro Education* 11:158–164, April, 1942.

2) Bernice Barnard, *An Experiment in Testing Music Capacity*. Unpublished Master's thesis, University of Idaho, Moscow, 1932, 18 pp.

3) Sylvia F. Bienstock, "Review of Studies on Music Aptitude," *Journal of Educational Psychology* 33:427–42, September, 1942.

4) Samuel T. Burns, "The Value of Prognostic Tests for Instrumental Pupils," *School Music* 31:6–9, March, 1931.

5) John E. Chadwick, *The Prediction of Success in Music*. Unpublished Master's thesis, Colorado State College of Education, Greeley, 1931, 50 pp.

6) L. W. Chidester and A. M. Small, "Preparatory Instru-

ments—Some Scientific Aspects," *Music Educator's Journal* 26:25, March, 1940.

7) LeRoy W. Darling, *A Study in the Correlation Between Achievement in Music and Academic Studies*, Unpublished Master's thesis, University of Idaho, Moscow, 1941, 15 pp.

8) Raleigh M. Drake, "The Relation of Musical Talent to Intelligence and Success in School," *Journal of Musicology* 2:38–44, May, 1940.

9) Paul R. Farnsworth, "Ratings in Music, Art, and Abnormality in the First Four Grades," *Journal of Psychology* 6:89–94, June, 1938.

10) Robert Gaines, "Guidance in Instrumental Music," *Music Educator's Journal* 29:21, January, 1943.

11) Josephine Harland, *An Experiment in Correlation Between Native Ability and Achievement in Music*, Unpublished Master's thesis, University of Idaho, Moscow, 1943, 23 pp.

12) Hugh Hartshorne, Mark A. May, and F. K. Shuttleworth, *Studies in the Organization of Character*, Vol. III, The Macmillan Company, New York, 1930, 503 pp.

13) Raymond K. Harris, *A Plan for the Testing and Measuring of Musical Talent Relative to the Beginning Band*, Unpublished Master's thesis, University of Idaho, Moscow, 1943, 23 pp.

14) Kenneth Hayes, *Guidance Factors in the Selection of Students for the Study of Instrumental Music in the Public Schools of Tulsa, Oklahoma*, Unpublished Master's thesis, North Texas State Teacher's College, Denton.

15) Mark Hindsley, "The Pre-Class Clinic," *Educational Music Magazine* 17:29, January-February, 1938.

16) Charles J. Lamp and Noel Keys, "Can Aptitude for Specific Musical Instruments Be Predicted?" *Journal of Educational Psychology* 26:587–596, November, 1935.

17) Thurber H. Madison, "Interval Discrimination as a Measure of Musical Aptitude," *Archives of Psychology* No. 268, 1942, 100 pp.

18) Harold C. Manor, *The Measurement of Persistence for Guidance Purposes*, Indiana University, 1946, 35 pp. typed.

19) James L. Mursell, "Measuring Musical Ability and Achievement: A Study of the Correlations of Seashore Test Scores and Other Variables," *Journal of Educational Research* 25:116–126, February, 1932.

20) Charles B. Neal, "The Musical Achievement of a Tonette Class in the Fourth Grade of the Zeigler, Illinois School," *Lyons Band News* 6:13, October, 1940.

21) Don P. Pence, *A Study to Determine the Optimum Grade for Beginning Class Study of Wind Instruments*, Unpublished Master's thesis, University of Kansas, Lawrence, 1942, 73 pp.

22) William D. Revelli, "Instrumental Adaption and Aptitude Tests," *Etude* 59:601–602, September, 1941.

23) David G. Ryans, "An Experimental Attempt to Analyze Persistent Behavior: I, Measuring Traits Presumed to Involve Persistence," *Journal of General Psychology* 19:333–335, October, 1938.

24) James H. Taylor, "Relationships of Finger Length and Hand Width to Music Ability," *Journal of Applied Psychology* 20:347–352, June, 1936.

25) J. G. Watkins, *Objective Measurement of Instrumental Performance*, Teacher's College, Contributions to Education, No. 860, Columbia University, New York, 1942, 88 pp.

THE RELATIONSHIP OF READING AND SPEECH DIFFICULTIES

THOMAS H. EAMES
School of Education, Boston University

There is an anatomical as well as a physiological or psychological relationship between speech and reading. Both are language functions, commonly located on the left side of the brain; the latter being centered in the angular gyrus (Brodmann area 39) and the former in the inferior frontal gyrus (Broca's area or Brodmann 44) together with parts of areas 4 and 6 in the precentral gyrus where the control of the muscles of speech are located. All of these, in common with the other centers in the language area are closely interconnected by association paths and with the less well localized psychic centers. The association paths between the speech and reading centers are of particular importance in oral reading, just as those between the speech and auditory areas (41 and 42 and probably part of 22 and 37) are significant in their relationship to the formation of speech through association of the sounds of words with the muscular movements required to produce them.

Defects, deficiencies, trauma or pathological processes in the various centers themselves give rise to well-known aphasias, both partial and complete, but interference with the association tracts between the centers may impede or block certain aspects of learning. It is well known, for example, that trouble in the auditory area may impair the ability to learn speech sounds and thereby interfere with their reproduction by the pupil. Difficulties in the pathways connecting the visual, auditory, and reading areas may interfere with the transmission of adequate stimulus impulses from one to another and impair the association of visual and auditory symbols. Similarly impairment of the association pathways between these centers and the speech area is likely to interfere with visual and auditory associations with speech.

The literature exhibits a growing interest in the relationship between speech and reading difficulties. Bennett[1] stated that children presenting a history of speech defects seem liable to reading failure, while [illegible] reading cases were found to exhibit marked defects of speech. Eames[3] reported five per cent more speech defects among reading failures than among non-failures. On the other hand Moore[4] studied a group of high-school pupils and found only a few cases of deficient reading among those exhibiting speech handicaps. A small positive correlation between speech age and reading age was reported by Davis,[5] while Bond[6] found no differences in the incidence of speech defectiveness among either good or poor readers.

Quite a number of investigators, like Gaines,[7] have pointed out probable relationships between deficiencies in reading and speech. For example, in studying first- and second-grade pupils, Rossignol[8] found that reading performance varied significantly with speech production. Generally, however, they appear to believe that factors which affect speech may also interfere with reading. Robinson[9] stated that dyslalia was the commonest cause of reading failure, and stuttering not at all so among thirty cases in her comprehensive study. She concluded that inadequate auditory memory span for sounds and insufficient auditory discrimination might be causes of both reading failure and speech difficulties in some cases. Monroe[10] made a similarly cautious deduction in one of her studies, indicating that speech difficulty might be a cause of reading failure or that both speech and reading troubles might arise from a common cause. Moser[11] explained stuttering as being due to a neuromuscular derangement which might also account for difficulty in reading. The frequent co-existence of speech defects and poor eye muscle control was pointed out by this investigator. Witty & Kopel[12] felt that emotional reactions to defective speech might contribute to reading and language failures, while Cobb[13] considered uncertain dominance an important cause of both reading and speech defects, although he did not regard strong left-handedness as being so. Bond,[14] Hildreth,[15] Monroe,[16] Robinson,[17] and Schonell[18] were in some agreement in giving considerable attention to weakness in auditory discrimination as a cause of both defective reading and defective speech. Davis[19] studied the relationship of oral reading to speech and came to the rather obvious conclusion that a child who articulates well is likely to read faster and more accurately than one who continues to use baby talk. Hall[20] gave attention to speech and silent reading, coming to the conclusion that there is no relationship between these functions. Of course, one would expect less relationship between speech and silent reading than between speech and oral reading. The anatomical connections in the brain support this. Travis[21] found stutterers retarded in the rate and comprehension of silent reading as well as oral, and Bloodstein,[22] investigating the reading rates of stutterers in the non-stuttering, fluent interval, found that stutterers read more slowly than non-stutterers in this period. This showed that, even when not stuttering, the stutterer is impeded in his reading by his language handicap.

Treatment of speech defective children continues to involve reading in some of its techniques. Choral reading is often employed to help remove a stutterer's emotional tension and to provide model speech which he hears and follows. Pattie and Knight[23] experimented with twelve subjects, having them read in six different situations. Their conclusion was that the removal of emotional tension is not a very important factor in the improvement of oral reading among stutterers during choral reading. However, they offered no definite evidence that hearing others speak facilitates the speech of the stuttering child. Heltman[24] advised choral reading with the stutterer being given solo lines.

There is no general agreement about the relationship between reading and speech failures or defectiveness. Probably this is due partly to some maturation or learning factor which produces different conditions at various age and grade levels. It is not unlikely that many of those who investigate the problem lack the breadth of training and experience to bridge the two fields well enough to explore their peripheral as well as central relationships. However, certain broad generalizations can be drawn.

1) Neurological lesions in the language centers or their interconnections may impair both speech and reading.

2) Failure or inadequacy of auditory association and discrimination may predispose to either speech or reading trouble.

3) Speech defects occur in a certain proportion of reading failures and *vice versa.*

4) Emotional reactions to speech difficulties may impair reading.

5) Oral reading is more difficult for a person with a speech defect.

The author's conclusion is that both speech and reading troubles are very likely to stem from the same basic defect, as may also other language function impairments. Essentially the problem is neuro-physiological with psychological overtones, but the inadequacy of present investigative methods make it necessary, at least for practical purposes, to subdivide the field and apportion the administration of treatment among persons who have made a special study of its various aspects. The weakness of this plan lies in the slowness of the various component groups to follow leads into adjacent language fields, due in large part to limitations imposed by economic, educational and demand factors. The strength of the plan lies in the immediate practical benefit made available to pupils with language difficulties by the person who knows something beneficial to do regardless of whether or not it is consistent with the optimum over-all understanding of language defects and deficiencies. Probably a better and more unified approach to the treatment of combined language difficulties will emerge with further research.

REFERENCES

1) C. C. Bennett. *An Inquiry into the Genesis of Poor Reading.* Contributions to Education 775, Teachers College, N.Y., 1938.

2) E. H. Stullken. "Retardation in Reading and the Problem Boy in School." *Elem. Eng. Rev.*, 14, May, 1937.

3) T. H. Eames. "Incidence of Diseases among Reading Failures and Non-Failures." *J. Ped.* 33: November, 1948.

4) E. A. Moore. "Reading and Arithmetic Ability Associated with Speech Defects." *J. Speech Disord.* 12: March, 1947.

5) I. P. Davis. "The Speech Aspects of Reading Readiness." *Nat. Elem. Prin.* 17: July, 1938.

6) G. L. Bond. *Auditory and Speech Characteristics of Poor Readers.* Contribution to Education 657, Teachers College, N.Y., 1935.

7) F. R. Gaines. "Interrelations of Speech and Reading Disabilities." *Elem. Sch. J.* XLV, April, 1941.

8) L. J. Rossignol. *The Relationships Among Hearing Acuity, Speech Production and Reading Performance in Grades 1A, 1B and 2A.* Contributions to Education 936, Teachers College, N.Y., 1948.

9) H. M. Robinson. *Why Pupils Fail in Reading.* Univ. of Chicago Press, Chicago, 1946.

10) M. Monroe. "Reading Aptitude Tests in Beginning Reading." *Education,* LVI: September, 1935.

11) H. M. Moser. "A Qualitative Analysis of Eye Movements During Stuttering." *J. Speech Disord.* 3: September, 1938.

12) P. Witty and D. Kopel. *Reading and the Educative Process.* Ginn & Co., Boston, 1939.

13) S. Cobb. *Borderlands of Psychiatry.* Harvard University Press, Cambridge, 1943.

14) G. L. Bond. *Op. Cit.*

15) G. Hildreth. "Speech Defects and Reading Disability." *Elem. Sch. J.* XLVI, February, 1946.

16) Monroe. *Op. Cit.*

17) H. M. Robinson. *Op. Cit.*

18) F. J. Schonell. *Backwardness in the Basic Subjects.* Oliver & Boyd, London, 1942.

19) I. P. Davis. *Op. Cit.*

20) M. E. Hall. "Auditory Factors in Functional Articulatory Speech Defects." *J. Exp. Educ.* 7, December, 1938.

21) L. E. Travis. *Speech Pathology.* D. Appleton & Co., N.Y., 1931.

22) O. N. Bloodstein. "The Relationship between Oral Reading Rate and the Severity of Stuttering." *J. Speech Disord.* 9, June, 1944.

23) F. A. Pattie and B. B. Knight. "Reading Difficulty Differences of Health Knowledge Tests." *Research Q.* 16: Am. Assn. Health & Phys. Educ. & Recreat. Oct. 1945.

24) H. J. Heltman. "Remedial Training for Speech Deviates in the Elementary School." *Elem. Sch. J.* XLVI, January, 1946.

A NOTE ON THE VALIDITY OF THE MENTAL AGE CONCEPT

RUKMINI S. RAMASESHAN
State University of Iowa*

The question raised here about the validity of the Mental Age is concerned with the generality of this unit of measurement. For example, is an MA of 8 earned by six-year-old and ten-year-old children the same MA? An MA score is an average of scores made on various types of tests involving a variety of mental processes. Since bright and dull children of a given MA earn the same average score, any differences among the subscores as between these bright and dull children would have to be compensating.

The purpose of this investigation is to test the validity of the Mental Age concept by determining whether or not there are systematic qualitative differences between bright and dull children of the same MA's on the subscores on certain tests. The tests used in the preliminary investigation are *The Chicago Tests of Primary Mental Abilities* and *The Iowa Tests of Educational Development.* These particular tests were analyzed for the purpose stated above, partly because the scores on approximately six hundred ninth-grade pupils were made available to the writer, by Dr. D. C. Shaw. In addition, the *Primary Mental Abilities* tests have the merit of having subtests of pure form and satisfactory reliability.

The two tests were administered to the entire ninth-grade population of two midwestern high schools. The task was then to select groups of students falling within given mental age categories on the *Primary Mental Abilities* tests and to form subgroups of bright (younger) and dull (older) pupils and to compare their scores on the subtests of this battery. The performance of the same subgroups of pupils was also compared on each of the subtests of the *Iowa Tests of Educational Development.*

The sampling of six hundred pupils from a single grade, the ninth, did not yield an adequate number of cases of bright and dull pupils having common MA's. However, the number appears to be sufficient to make the report of the results worth while. Had the six hundred tests been administered to pupils spread over two or three grades, instead of one, more useable cases would, of course, have been available.

---

* This article is a partial report on the writer's Master's Thesis. The work was done under the direction of Professor J. B. Stroud.

The mean MA of the entire sampling was 186 plus months, on the *Primary Mental Abilities* tests. An MA category of seven months, 183.5–189.4, was determined. The pupils in this MA group whose chronological ages fell above 177.7 (the mean CA of the pupils in this MA category was 171.7) were placed in the older or 'dull' group; those whose CA's fell below 165.7 were placed in the younger or 'bright' group. Thus the minimum difference in CA between the so-called dull and the bright pupils is one year of CA. It will be observed that dull here merely means a lower position in the group. Actually the mean IQ of the 'dull' group is approximately 100. Two seven month MA categories were likewise established above MA 189.4 and two below MA 183.5, and bright and dull subgroups similarly formed. Altogether these groupings yielded fifty-three bright and thirty-five dull pupils. The comparative scores of these two groups on the respective subtests of the *Primary Mental Abilities* tests are as follows:

	Number	Verbal Meaning	Space	Word Fluency	Reasoning	Memory
Bright Pupils (Mean IQ 115.2)	198.45	185.54	165.50	176.71	207.07	180.89
Dull Pupils (Mean IQ 101.5)	202.01	174.03	177.30	181.34	191.85	185.91

The mean score of the bright group on all six tests is 185.69; that of the dull group 185.51.

The bright group is significantly superior to the dull group on Verbal Meaning and Reasoning, the differences being 5.93 and 7.21 times, respectively, the SE of the differences. The dull group is significantly superior to the bright group in space, the difference being 3.49 the SE of the difference. On Word Fluency a difference was obtained of 2.63 times the SE of difference favoring the dull group.

The bright group exceeded the performance of the dull group on all the subtests of the *Iowa Tests of Educational Development.* The differences are significant in the case of only three subtests:

Correctness in Writing	(Dif./SEdif. = 5.20)
Quantitative Thinking	(Dif./SEdif. = 3.47)
General Vocabulary	(Dif./SEdif. = 3.07)

The foregoing results, while limited in scope, suggest that there are systematic qualitative differences between bright and dull children of the same MA and that the validity, meaning generality, of the Mental Age concept is open to some questions. Between children widely separated in Chronological Age qualitative differences in MA may well be fairly large and carry with them different meaning so far as success in school is concerned. However, the latter point is principally yet to be determined, since, in addition to the smallness of the sample used in this investigation, some of the subtests of the *Primary Mental Abilities* tests do not correlate highly with scholastic achievement.

BOOK REVIEWS

LAWRENCE A. AVERILL. *The Psychology of the Elementary School Child.* New York: Longmans, Green and Co., 1949, pp. 459.

This book is addressed to teachers of elementary school children rather than to the student of child psychology. It will probably find greater use in departments of education engaged in teacher-training than in departments of psychology or child development. It covers chiefly the child of elementary school age, although there is one chapter, 16, rather a minor one, devoted to Early Childhood Training.

The book lives up to its title. It treats the psychology of the elementary school child, not the psychology of teaching the elementary school subjects. Additional chapters are as follows:

Personality Adjustment of the School Child
The Emotions of the School Child
Attitudes and Habits
Physical and Motor Growth and Development
Play Interests of Children
Social Development of Children
Language and Speech
Imagination and Make-believe
Juvenile Aesthetics
Maturation and the Motivation of Learning
Intelligence and Individual Differences
Meaning and Perception
How Children Learn
How Children Think
Guidance

The author seems to reflect the training he received at Clark in what may be called the classical period of Hall and Burnham. The book is stylishly written and on the whole presents the teacher-in-training with a pretty good description of what elementary school children are like, psychologically. It has occasionally occurred to the writer that the children pictured are ideal children or otherwise a bit unusual, although this impression may well result from the author's style of writing. The book presents the child of school age in action. It is written out of a wide experience. Since there is not a great deal of documentary or experimental evidence presented, an instructor with limited background might find the book a little difficult to teach. It seems that students will enjoy reading it. J. B. STROUD

State University of Iowa

G. W. PARKYN. *Children of High Intelligence.* Auckland, N. Z.: New Zealand Council for Educational Research, 1948, pp. 228.

This volume is a presentation of the data from a study designed to provide information about highly intelligent children in New Zealand.

The author briefly discusses the meaning and importance of high intelligence drawing heavily upon the work of Burt, Spearman, and Hollingsworth in the formulation of his concepts. He also reviews the studies by Terman and others in the area of characteristics of gifted or highly intelligent children. He prefers the term high intelligence, reserving gifted for individuals with specific talents.

The data in the six-year intensive and extensive survey was based on three groups of children in the upper five per cent of the population as determined by standardized intelligence tests. The two groups selected by means of the Otis Self-Administering Test of Mental Ability were used for statistical treatment while detailed case study methods were applied to a smaller group selected by means of the Revised Stanford Binet Scale 1937. Within the limits of the instruments and methods used, the study was carefully carried out and cautiously analyzed with recognition of several possible interpretations.

The conclusions indicate that with the exception of differences resulting from local cultural and social influences, the characteristics of children of high intelligence in New Zealand are comparable to those of children studied in the United States and England.

The remainder of the book is devoted to stating the problems involved in the education of highly intelligent children, including learning, teaching methods, curriculum and ability grouping. He concisely gives the advantages and disadvantages of the various approaches used in other countries, the present status in New Zealand, and suggestions for reorganization.

While the author is open-minded and strongly advocates experimentation to determine the best methods for educating children of high intelligence, his present point of view is rather conservative.

There is little that is new in the book. It's contribution is in its resumé of previous studies and its verification of earlier data, in the fair presentation of contradictory points of view, and in the summary of the problems.

GWEN F. ARNOLD

The University of Wisconsin

CHARLES MORRIS. *The Open Self.* New York: Prentice Hall, 1942, pp. 174.

This is a book with a thesis and moral lesson. In essence, the sermon is: "Diverse open selves united on the common ideal of an open society of open selves can defend, celebrate, and enhance their own and each other's uniqueness." This doctrine of open selves in an open society is presented by Morris in a very convincing manner, and in his presentation he reveals a familiarity with philosophy, psychology, sociology, as well as expresses an interest in the welfare of mankind and the destiny of man. His sense of values is fundamentally sound and matured; his direction and method for attaining goals in society are well expressed, and the book is well written under such popular captions as "Man as His Own Maker," "Man's Knowledge of Man," "A Primer of Semantics," "Ways to Live," "The Catacombs of the Self," "The Universe of the Many," "The Open Society of Open Selves," and "Freedom or Frustration."

Of interest to educational and social psychologists is the author's report of a survey of a thousand college students giving their reaction to thirteen possible ways of life, with the conclusion that forty per cent choose dynamic integration and diversity. The thirteen ways to live are identified with philosophical statements and quotes from literature well known. Illustrative is Path Number One. Here the individual actively participates in the social life of his community not to change it primarily, but to understand, appreciate and preserve the best that man has obtained. Another path, Number Four, is that life is something to be sensuously enjoyed, enjoyed with relish and abandonment. Still another is the use of the body's energy as the secret of a rewarding life. The conclusion is that most people favor diversity, and the next largest number—seventeen per cent—favor nothing in excess. Another ten per cent favored careful, wholesome enjoyment.

The doctrine of the open self and open society is offered by Morris as the only solution of man's ills and anxieties of the present time. This he presents as the most preferred road to freedom, because, as he sees it, the totalitarian society flourishes on the food of anxiety, is temporarily effective through its promise to relieve anxiety, but is ultimately unsatisfactory because it lives only by the perpetuation of anxieties. His diagnosis and prescription is relatively simply expressed; it is merely that the danger of America is not senility but sterility, that the new American frontier is human rather than geographical, and that the new American frontier marks our responsibility and our opportunity. In his actual prescription he reveals the fundamental soundness of his philosophically-determined goals and also reveals one obstacle toward clear understanding that runs through the book; namely, identifying of personality traits from Horney's writings of dominance, submission and dependence with Sheldon's concepts endomorph, mesomorph, and ectomorph. In his prescription he presents it in this fashion: "We need our endomorphs to keep us humanly warm, our ectomorphs to keep us humanly sensitive, and our mesomorphs to keep us humanly bold and adventuresome, and on this frontier we can advance together. Here we can use all of ourselves." All is well, all is right and all is in line with liberated thinking of the day. What is unconvincing and somewhat irrelevant is his hook-up of human values with Sheldon's concepts loosely considered. If Morris never knew anything about endomorphs, ectomorphs and mesomorphs, he would have written an even more clearly, more intelligible book, more understandable by more people for whom the volume is written. H. MELTZER

Psychological Service Center
St. Louis, Missouri

JEAN PAUL SARTRE. *The Psychology of Imagination.* New York: Philosophical Library, 1948, pp. 285.

An existentialist takes an excursion into the phenomenology of imagination and the imaginary in this book called *The Psychology* *of Imagination* by Jean Paul Sartre. What the author says he is trying to do is "describe the great function of consciousness to create a world of unrealities of 'imagination' and its noetic correlative, the 'imaginary.'" What he calls the image family is considered by him in Part I under the caption of "The Certain." This includes a consideration of such topics as the sign and the portrait, the consciousness of imitations, schematic drawings, faces in the fire, spots on the wall, rocks in human form, hypnagogic images and persons seen in coffee-grounds and a crystal ball. Part II is entitled "The Probable." Here are considered such items as knowledge, affectivity, movements and the rôle of the world in a mental image and how the things appear in a mental image. Part III, called "The Rôle of the Image in Mental Life," includes a consideration of the symbolic schemes, and Part IV, "The Imaginary Life," includes a consideration of image and perception, the unreal behavior, pathology of the imagination and the dream. The concluding chapters consider consciousness and imagination and the work of art. The book contains an index.

The image and the perception are considered by Sartre as two elementary psychical factors of similar quality which represent the two main irreducible attitudes of consciousness. The easiest way to reveal the nature of the content found in this book is through illustrative quotes:

"For an image is not purely and simply the world negated, it is always the world negated from a certain point of view, namely, the one that permits the positing of the absence or the non-existence of the object present 'as an image.'" (p. 268). "Nothingness can present itself only as an infra-structure of something. The experience of nothingness is not, strictly speaking, an indirect one, it is an experience which is 'in' principle given 'with.'" (p. 271). "The unreal is produced outside of the world by a consciousness which stays in the world and it is because he is transcendentally free that men can imagine." (p. 271). "The values of the Good presume being in-the-world, they concern action in the real and are subject from the outset to the basic absurdity of existence. To say that we assume the attitude of esthetic attitude to life is to constantly confuse the real with the imaginary." (p. 281).

The foregoing illustrated quotes reveal the style and the intellectual excursion the reader would take with the existentialist and novelist who himself takes an excursion into what he thinks is the psychology of imagination.

To Sartre's consideration of the negative one is tempted to react with Morris Bishop's short poem in the recent issue of the *New Yorker* called "Not Unmindful of the Negative as I am Not" which ends with the line, "If Not's not not, then prithee, what's Not not?"

The Chips are Down is the name of a short novel by the author and well describes the attitude with which persons who are serious students of psychology without an interest in side-track reading are likely to emerge if what they expect out of this volume is a factual consideration of what psychology knows about imagination. *The Chips are Down* for such readers. Only if they are willing to take an excursion with Sartre's intellectualizing about imagination will they get any enjoyment out of the book.

H. Meltzer

Psychological Service Center
St. Louis, Missouri

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 41 February, 1950 Number 2

STUDIES OF THE GROUP RORSCHACH IN RELATION TO SUCCESS IN THE COLLEGE OF THE UNIVERSITY OF CHICAGO*

LEE J. CRONBACH
Bureau of Research and Service
College of Education, University of Illinois

Widespread interest has been aroused by Munroe's monograph[4] on the adjustment of college students and their scholastic success, in relation to predictions based on the Rorschach test. Munroe found that even a rapid scoring of the Group Rorschach yields a useful predictor of academic success, and that the test draws attention to students likely to need help in emotional adjustment. The Munroe study leaves need for further research, however. Sarah Lawrence College, with its methods of adapting the curriculum to the individual, is sufficiently atypical that findings there may not apply elsewhere. Further studies are also needed to obtain a clearer indication as to what aspects in the Rorschach performance contribute to the predictions.

Many colleges have planned studies along the lines of the Munroe work. Some reports, generally supporting the claim that Rorschach data have predictive value, have appeared.[3,6,7] Some of this evidence, however, is weakened by faulty statistical reasoning.[1] The present study was begun as a direct repetition of the Munroe study, to predict academic success at the University of Chicago. As the study proceeded, the mass of data collected made several other studies possible. The present report is a statistical threshing of the material available. Clinical analysis is also planned at a later date.

---

* This study was planned and the data were assembled by the University Examiner's Office of the University of Chicago. Contributions to the study were made by Benjamin S. Bloom, George De Vos, Esther A. Frankle, and Harriet Moore. The present writer is primarily responsible for the analysis and interpretation.

Basic data.—The College of the University of Chicago is the top unit in a 6-4-4 system, comprising grades XI–XIV of the usual ladder. According to the Chicago plan, students take general courses in several major areas, the work being presented principally through small discussion classes supplemented by a weekly lecture. The student is much less closely supervised than he would be in the usual high school. His official record of progress is based exclusively on comprehensive examinations given at the end of the school year, rather than on such bases as class attendance, teacher marks, and teacher-made tests given throughout the year. These tests are prepared, given, and marked by a central examining staff. While the testers work closely with the instructors in preparing tests, the final form of the examination is prepared by the examiner and the final grading is made without knowledge of the identity of the writer of each paper, and thus is free from certain types of bias. This practice provides an especially interesting and reliable criterion for a study of the Rorschach as a predictor. The examinations emphasize comprehension of principles and development of intellectual skills, rather than recall of isolated facts.

In 1946, a group of three hundred students entering the College were given the Group Rorschach in the customary manner, and the American Council Psychological Examination. At the end of the year the comprehensive examinations in the Natural Sciences, Humanities, Foreign Languages, Social Sciences, Mathematics, and English were given. Subjects in the present study include only those students who took from three to five comprehensives in the first year; the grades were averaged, with equal weighting, to yield the grade point average which is the principal criterion in this study.

The subjects tested are not representative of the usual 'college freshman.' Most are from homes which can afford a substantial tuition. They have passed severe entrance hurdles. They are young, the median age being about fifteen. They are intellectually extremely superior; on the ACE test, total score, approximately fifty-three per cent surpass the ninetieth percentile of the national norms for thirteenth-graders. The group undoubtedly have special motivations which cause them to seek out the University of Chicago program.

Problems investigated.—The following questions are considered in this report:

(1) What are the characteristics of these students in terms of the major categories?

(2) How adequately does adjustment judged from the Rorschach predict success at Chicago? Does it enhance prediction based on the ACE?

(3) What signs in the Munroe check list are associated with academic overachievement or underachievement?

(4) What signs are correlated with the ACE score, and with the difference between Q and L? The latter question was raised by Munroe's study[5] that links the pattern of abilities to personality.

(5) How does Rorschach performance relate to ratings of social adjustment and to personal effectiveness as measured by sociometric techniques?

It is the purpose of all these questions to identify the possible usefulness of the Group Rorschach in college personnel work.

Analysis of the data.—Because of the intensive analysis planned, a thorough scoring of the Rorschach was made. A first scorer had marked the check list in a very rapid fashion, and treatment of the results showed that the data did not predict marks. But it was apparent that some responses which should have led to entries on the check list had been overlooked in the hasty scoring. In a rescoring by one of two new scorers, each Rorschach response was scored. The check list was then marked carefully by the scorer, and in many cases the checking was reviewed by the other scorer. The data may have been analyzed more intensively than Munroe's records, which would lead to somewhat higher check-list scores, on the average, for this group.* The second scorers also assigned a letter grade indicating estimated adjustment, using Munroe's scale,[4] (p. 21) which ranges from A (sound integration) to E (psychopathology).

Contamination was eliminated by the facts that Rorschach findings were not made available to instructors or examiners, and that scorers marked the tests without knowing the students and without referring to non-test data other than age and sex.

---

* In retrospect after completing the study, it appears that the rapid scoring gave neither better nor worse predictions than the more careful check-list scoring.

RESULTS

Because time was not available for rescoring all records, only two hundred cases—a sample of the three hundred persons tested—were used. These are a random sample except that all seventy-two of the cases residing in college dormitories were included. Table I summarizes the distribution of cases. There were, in all, seventy-one men and one hundred twenty-nine women. In some analyses, the first-year group was treated separately from the advanced students.

TABLE I.—SUBJECTS CLASSIFIED BY AGE, SEX, AND GRADE PLACEMENT (N = 200)

Age	Placement at Time of Entrance								Total
	First Year (11th grade)		Second Year (12th grade)		Third Year (13th grade)		Fourth Year (14th grade)		
	Men	Women	Men	Women	Men	Women	Men	Women	
Below 15	4	2	0	0	1	0	0	0	7
15-0 to 15-11	29	21	3	2	1	0	0	0	56
16-0 to 16-11	24	16	7	7	0	0	0	0	54
17-0 to 17-11	1	1	6	8	6	6	0	0	28
18-0 to 18-11	0	1	0	0	5	3	0	0	9
19-0 to 19-11	2	0	0	0	1	2	0	0	5
20-0 Over	2	0	0	0	29	2	8	0	41
Total	62	41	16	17	43	13	8	0	200

RORSCHACH CHARACTERISTICS OF THE SAMPLE

Our first question is what the Chicago group is like, as reflected in Rorschach performance. One way of looking at this is to see how frequently each of the Munroe 'checks' occurs, since a check is allowed for each variation in personality which may suggest adjustment difficulty. This tabulation is not so rich as a clinical analysis of each record would be, but it is relatively objective. The incidence of signs* in the Chicago group was as follows:

---

* These signs are used as defined by Munroe[4] (pp. 87–101).

Over 50 per cent of the subjects showed: shading shock (83 per cent), color shock (92 per cent; two checks given to 57 per cent), FC− (56 per cent), C:M−(66 per cent).

31–50 per cent of the subjects received one or more checks for: S+, AtSx+, F(B or V), FKFc−, m+, Total Movement+, Total color−.

20–30 per cent of the subjects received one or more checks for: W+, Suc r, Range−, F per cent+, M+, M−, CF+.

Munroe has provided us with the incidence of signs in one hundred twenty Sarah Lawrence freshmen (girls), entering in 1945–1947. The general distribution of signs resembles that at Chicago, with certain marked exceptions. The following signs were noted by Munroe in at least ten per cent more cases than at Chicago: CF−, Range−, W+, Total mvt.+. The Chicago scorers noted these signs at least ten per cent more often: C+, Suc r, F(B), F(V), AtSx+, Total color−, S+, FKFc−, m+, FC−, C:M−, shading shock, color shock. The most striking differences in incidence relate to shading shock (Chicago 83 per cent, Sarah Lawrence 35 per cent), color shock (92 per cent, 60 per cent), FC− (56 per cent, 28 per cent), m+ (46 per cent, 14 per cent), FKFc− (37 per cent, 7 per cent), and Suc r (27 per cent, 0 per cent).

One cannot be sure that in a group having exceptional scholastic aptitude the signs of maladjustment are the same as those for persons representing a broader range of ability. Moreover, the differences in age and sex between the two samples are too great to permit confident interpretation of the comparison.

If the conventional Rorschach interpretations hold for this group, the Chicago students show a larger number than usual of signs of emotional disturbance (shading and color shock, F(BV), AtSx, m), inflexibility (S, range, F per cent), and inward rather than outward orientation (M, color). The principal quantitatively-scored difference from the Sarah Lawrence group appears to be in the color balance—Chicagoans tending relatively more toward C and CF than FC—and in the tensions represented by m. If the scoring in the two schools on more qualitative items such as color shock can be compared, the Chicagoans are also less flexible and more prone to emotion in threatening situations.

PREDICTION OF GRADE AVERAGE

Does the Rorschach improve prediction of grades beyond that obtained from the ACE Test? The answer given by our data is almost completely negative. The correlations of grade average with several variables are as follows:

with ACE Total	.45 (first year), .39 (2d–4th year)
with number of Rorschach checks	.17 (first year), .00 (other years)
with Rorschach adjustment rating	.25 (first year), .07 (other years)

Positive correlations show that good adjustment is associated with good grades. For one hundred cases, a product-moment r of .00 has a standard error of .10. The correlation of Rorschach grade with ACE Total is .11 (first year), .29 (other years). A multiple R computed from ACE and Rorschach grade is .49

TABLE II.—RORSCHACH ADJUSTMENT RATING COMBINED WITH ACE IN RELATION TO ACADEMIC STANDING (FIRST-YEAR STUDENTS)

Rorschach	High ACE (22* and over)				Low ACE (Below 22)				Grand Total
	A or B	C	D or E	Total	A or B	C	D or E	Total	
Grade average†									
3.0–3.9	8	9	1	18	6	4	1	11	29
2.0–2.9	6	9	2	17	4	15	8	27	44
1.0–1.9	1	2	1	4	4	11	5	20	24
0.0–0.9	1			1	1	1	0	2	3
Total number	16	20	4	40	15	31	14	60	100‡
Percentage earning 3.0 and above	50	45	25	45	40	13	7	18	29
Percentage falling below 2.0	12	10	25	13	33	39	36	37	27

* On a standard-score scale with mean 20, s.d. 4.

† 3.0 equals B.

‡ ACE scores were missing for three cases.

(first year), .39 (other years). The improvement over prediction from ACE alone is negligible.

Table II, corresponding to Munroe's Table VII[4] (p. 49), shows the grade distribution for each combination of ACE and Rorschach standings for first-year students. There is some relationship between the Rorschach and grades earned. Bright students tend not to earn high grades, unless they have at least a fair degree of adjustment. With low ACE score, students' chances of receiving high grades rise or fall in proportion to their adjustment. The trend is insufficient to affect the multiple correlation appreciably. These sub-hypotheses regarding the interaction of Rorschach and ACE scores might prove to be significant in a larger sample of cases.

CHARACTERISTICS FAVORING ACHIEVEMENT

Even though the total Rorschach score does not predict achievement, there may be some traits revealed by the Rorschach which do correlate with academic success. The separate signs in the Munroe list were therefore correlated with tendency to underachieve. To do this, the scatter diagram between ACE Total and Grade Average (for all three hundred cases) was divided into two parts. The twenty-fifth percentile was found in each column representing an ACE score-interval, and a smooth line was drawn through these points. In the sample of one hundred ninety-seven cases for whom we had ACE scores and rescored Rorschachs, the line cut off sixty-eight underachievers. One hundred twenty-nine cases fell in the normal-to-superior group. This division permits comparison of students who differ in achievement with ACE score held constant. The failure to divide the group in the ratio 25:75 reflects smoothing and grouping, rather than differences between the rescored group and the parent group.

The two subgroups were compared on all the check-list entries, and on a few other scores. Relationships were generally negligible. The relations in Table III are reported only as leads for further research. What trends there were seemed to be linear; no sign is characteristic of both overachievers and underachievers to a greater degree than normal achievers. Not a single one of the forty-four relationships reaches significance by the chi-square test, both variables being dichotomized and

Yates' correction being used. Virtually none of these signs are confirmed by Thompson[7] as having a relation with grades, ACE score being held constant. There is no basis in the present study for advising further attempts to predict marks at Chicago by statistical treatment of Rorschach check-list entries.

TABLE III.—RORSCHACH SIGNS HAVING POSITIVE BUT NON-SIGNIFICANT RELATIONS WITH UNDERACHIEVEMENT

Characteristic	Percentage Incidence		r_{tet}
	Over-achievers and Normal Achievers (N = 129)	Under-Achievers (N = 68)	
W− −−	10	19	.28
c + ++ +++	19	10	.25
Tot Movement + ++ +++	43	28	.24
FM + ++ +++	23	10	.22*
Dd + ++ +++	14	7	.18

* Product-moment *r*.

CORRELATION WITH ACE SCORE

Because there has been much interest in the possible relation between Rorschach performance and tested intelligence, a comparison was made between the ACE total score and each of the check-list signs. In this high-level group, with restricted range, tetrachoric r's were all small, save that FKFc− correlated −.45 with ACE total. Chi-square, corrected, for FKFc was 9.3 (but this was only one out of twenty-five significance tests in this section of the study, and is therefore barely significant if at all). Other non-significant correlations were Dd+, .40; FC'−, −.30; C'+, .30, FM+, −.25; m+, .25; M+, .20. These associations are so low as to discourage attempts to make a statistical estimate of psychometric intelligence from the Rorschach signs in a group such as this. The unreliability of single-test signs would obscure any underlying relationship. High ACE score is associated with presence of FKFc, Dd, FC, C′, m, and M responses, and associated with infrequent use of FM. In a corresponding study with less homogeneous subjects, Thompson[7] found similar associations for M, FM, and m, and apparently for FC. Low Dd was found with high ACE in her group. Her scoring does not permit comparison of the two studies on FKFc and C′. On the other hand, she lists numerous correlations (significance not reported) of ACE with Rorschach signs which were found not to be differentiating in the Chicago group. Wittenborn's recent study[9] of a small sample suggests that few relationships between Rorschach scores and mental ability exist for college students. He finds slight association of tested ability with R, M, and C′.

THE Q-L DIFFERENTIAL

Munroe found numerous Rorschach differences between girls with L higher than Q and girls with Q higher than L on the ACE. Our data are in slightly different form from Munroe's, but permit further investigation of the same question. Because of the preselection of Chicago students, our cases come from a markedly different portion of the Q-L plane than hers.

Our Higher L group consisted of all students whose standard score in L was at least ½ s.d. higher than their Q score. The Higher Q group consists of those whose Q score was at least ½ s.d. higher than their L score. There were nine, twenty-seven, and thirty girls, respectively, in the Higher Q, equal, and Higher L groups. In the same categories, there were respectively thirty-seven, forty-two, and forty-four men. Because of this sex difference, we tabulated Rorschach performances of the two sex groups separately. For significance tests among the girls, the equal and Higher Q groups were thrown together. Among men, significance tests were based on chi-square applied to the Higher Q versus the Higher L group.

Absolutely no difference was found which could be considered statistically significant. On many items, small differences appeared in one sex but were accompanied by just the opposite difference in the other sex. Munroe reports five significant differences: Higher L is associated with high M per cent, high total movement, high F(BV), high C′, and low F per cent. Among our sample of girls, there was no difference in F(BV), M, and Tot mvt., a difference confirming Munroe in F per cent, and a difference contrary to Munroe's in C′. The tendency of Higher Q girls to show excess F per cent is not statistically significant, but it is in accord with previous reports by Wells[8] and Munroe. Among girls, the percentages receiving a check for overemphasis on F per cent were: Higher Q, 33; equal, 11; Higher L, 7. Among boys, the percentages were 43, 31, and 26.

In our sample there is no appreciable relation between Rorschach scores and the ability pattern. It may be that this is a result of special characteristics of the Chicago sample. Even the lowest of the Higher Q students at Chicago probably has considerable ability along linguistic lines, or he would not have reached the College. Furthermore, the Chicago program, which stresses general education, may not attract students with extreme overspecialization in either the linguistic or quantitative directions. Our results do definitely discourage generalization about "the type of personality likely to have Q greater than L," and *vice versa*, until further research has been done on subjects of many levels and many ages.

PREDICTION OF SOCIAL ADJUSTMENT

Munroe found that Rorschach scores permitted her to predict which students would later show signs of maladjustment, as judged by their use of psychiatric service and other evidence. No criterion available at Chicago is directly comparable to Munroe's. As one criterion of maladjustment, ratings from dormitory resident heads were obtained. For each first-year student living in a University residence hall (and this includes all those not residing at home in Chicago), the head of the dormitory unit in which he lived supplied a rating. Twelve raters were involved, each one rating from three to eleven students. No student was rated by two heads. The raters were professional people living in the dormitory and working with the students, but they had very limited psychological training. The rating was made on a five-point scale which paraphrased Munroe's scale used in assigning Rorschach letter ratings of adjustment.

The correlation of the Dormitory Rating with Rorschach letter rating is: for forty men, .38; for thirty-two women, .31; for both combined, .31. The first correlation is significant at the one per cent level, the second at the five per cent level; the null hypothesis may be rejected. In view of the doubtful reliability and validity of the criterion, the correlation encourages the belief that the Rorschach may prove useful in predicting social maladjustment in college. For number of checks, the correlations were: men, .27; women, .20; both, .24. Although we have no idea what to conclude, we must report the additional fact that for all cases combined the mere number of Rorschach responses correlates .33 with Dormitory Rating.

RELATION OF RORSCHACH RESULTS TO SOCIOMETRIC RATINGS

A second criterion of adjustment which may have considerable validity is a sociometric report obtained by Frankle.[2] Each resident of the dormitory was asked to rate each other resident of the same sex on a 'Guess Who,' or reputation, questionnaire. The questionnaire presented twenty bi-polar items; for example, "Which ones [in your dormitory] always seem cheerful and good-natured? Who are the ones who never look happy; always seem sad, worried, or frustrated?" Each person was provided with a list of all students of like sex residing in any dormitory, and was asked *to write after each question the name of any person on* the list who fitted the description. Respondents were allowed to name several persons for one question, or to omit names if the description fitted no one. While the students had been living together for a year, the dormitories were physically divided into 'entries,' so that each person knew a few students in his entry better than others in the group whom he saw only at mealtimes, in the lounge, and in general contact about the campus.

Among the twenty traits, there were five, such as 'shy versus show off,' which could not be considered as having one 'good' pole. Eliminating these five items, the other items could be considered as providing an index of adjustment, since mention in connection with one of the questions was obviously a favorable report, and the other question implied unfavorable adjustment. Eleven of the traits are listed, in condensed wording, in Table IV. The total number of mentions under favorable traits, minus the number under unfavorable traits, was taken as a Sociometric Rating of adjustment. This score may be taken as a net appraisal

TABLE IV.—RELATIONS BETWEEN TRAITS AS REPORTED IN REPUTATION QUESTIONNAIRE AND SELECTED RORSCHACH CHECK-LIST ENTRIES

Questionnaire Item	Associated with	Magnitude of Tetrachoric r	Check-list Entries Having r Less Than .20 with Questionnaire Item
Always seems worried (vs. cheerful)	Shading shock	.25	AtSx, F%, C', M
	Emphasis on Kk	.50	
Makes people feel comfortable (vs. makes people uncomfortable)	Emphasis on c	.40	M, Col sh, C:M, FC, FKFc, F%
Jumpy, jittery (vs. calm, unperturbed)	Emphasis on m	.40[a]	M
Always has something to gripe about (vs. handles problems adaptively)	Underemphasis on S	.40*	M, F%, F(BVE)
	Underemphasis on FKFc	.25	
	Underemphasis on FC	.55	
	Emphasis on CF	.50	
Overemotional, excitable (vs. cold and unmoved, overcontrols emotions)	Underemphasis on FC	.40*	M, F%
	Emphasis on CF	.35	
'Show off,' always trying to attract attention (vs. shy)	Emphasis on Kk	.20*	Tot C, C:M, FKFc
	Shading shock	.40[b]	
	Color shock	.30[b]	
	Emphasis on CF	.40	
Must be with other people (vs. prefers to be by self)			Tot C, C:M, F%
Acts like a 'high-school kid' (vs. mature)	Underemphasis on FM	.25*	M, CF, FKFc, A%
	Emphasis on FC	.35*	
	Emphasis on C	.40	
	High F%	.25*	
Interested in the abstract, theoretical (vs. concrete)			W
Spontaneous, flexible (vs. rigid, can't change views)	Form not BVE	.35	F%, M, CF
Withdrawn from reality, daydreams (vs. realistic attitude)	Emphasis on M	.05[c]	F%, F(BVE)
	Underemphasis on FM	.25*	

* Relationship is in direction opposite to that hypothesized.

[a] Lowest marginal total is only 12 per cent; when cut near middle of m continuum, r drops to .12.

[b] No advance hypothesis as to probable direction of this relation.

[c] Lowest marginal total is only 12 per cent; when cut is made near middle of M continuum, r drops to .05.

of the student by his peers. A large positive or large negative score is probably meaningful, although a score of zero could be achieved by a person with medium adjustment, by one whose good and bad mentions cancelled, or by one so inconspicuous that others were not acquainted with him and did not mention him. Because of the possibility that the number of 'mentions' might more nearly represent popularity than adjustment, the sociometric report was also scored by a weighted method, counting the number of traits on which the person had at least four more favorable than unfavorable mentions, minus the number of traits on which he has a net of four or more unfavorable mentions. This gave results no different from the first scoring procedure.

For men, the Sociometric Rating correlates .284 with Rorschach adjustment rating; for women, the correlation is .140; for both, .196. The correlations for men and for the total group are significant at the five per cent level. But for all cases combined, the number of Rorschach responses correlates .37 with the Sociometric Rating!

Many of the trait-descriptions in the reputation questionnaire employed terms often used in interpreting the meaning of specific Rorschach signs. For example, high F per cent is often thought to be associated with overcontrol. An examination of the correlation between F per cent and frequency of mention on the characteristic 'overcontrols emotions' is an unusually direct method of determining the validity of description from Rorschach scores. Each item in the Reputation Questionnaire was examined, and a list made of all the Rorschach check-list entries that might be expected to correlate with that item, in view of the meanings equated with Rorschach scores in the Rorschach literature. A tetrachoric correlation was then computed between the score of the student on the bi-polar item (number of favorable minus unfavorable mentions), and the frequency of the Rorschach check-list entry in question. Fifty-one hypotheses were tested in this manner.

This procedure is open to many criticisms. In the first place, analysis of scores one at a time violates the Rorschach maxim that each score must be interpreted in relation to the total pattern of the subject. One hypothesis set up was that subjects who were lacking FC would be rated 'cold and unmoved, under-controls emotions.' But some people scored FC− by Munroe's method are high in CF, whereas others who lack FC lack all color and are indeed 'cold' personalities. At best, the analysis presented here shows the extent to which single Rorschach check-list entries may be interpreted, and does not provide a basis for judging the validity of a clinician's interpretation of an entire record. If some hypotheses were tried which, on the basis of 'second guessing,' appear to have been unreasonable, it may be pointed out that the significance test employed does not penalize us for trying a large number of improbable hypotheses. The significance test employed introduces a penalty which lowers significance only when the results turn out to be directly opposite to the relationship hypothesized. Another serious difficulty is the extreme unreliability of tetrachoric correlations, especially when one variable is dichotomized so that a small proportion of the cases fall in one marginal total. No single correlation reported for this sample of seventy-two cases can be considered a precise expression of a relationship. Finally, one may question whether dichotomizing the check-list entries and the reputation-questionnaire entries is entirely satisfactory. On the trait 'show off-vs.-shy,' it might have been better to divide the cases into three groups, since some subjects showed neither of the polar traits. Among Rorschach scores, it might have been better to separate the subjects into three groups such as $M-$, M normal, $M+$, rather than dichotomizing. Too few cases were available for this type of analysis with our data.

We explored the possibility that different relationships between questionnaire items and Rorschach would be found for men and women, considered separately. Some differences were found, but the number of cases available to us are too small for our results to be worth reporting in detail.

Table IV reports all correlations investigated between Rorschach signs and reputation reports. Many correlations fell below the arbitrary critical limit of .20. Some of these failures to find relations are a contradiction of Rorschach theory. People with absence of Fc or with excessive F would be expected to make others uncomfortable; those with emphasis on W are presumed ordinarily to be interested in the abstract and theoretical. Disregarding two relationships where the direction of relation had not been postulated before studying the data, there were nineteen correlations greater than .20. Of the nineteen, twelve were in the direction anticipated, and seven were in the direction opposite to the hypothesis based on Rorschach theory. This is not a significant departure from the null hypothesis that fifty per cent of the correlations would fall in the predicted direction. (The standard error of a p of 50, for nineteen events, is .11). This analysis therefore failed to confirm the validity of inferences regarding traits from single Rorschach check-list entries. Many explanations can be advanced: the unreliability of single entries and the criterion, differences in the use of such words as 'mature' by student raters and Rorschach interpreters, etc. Even after these attempts to rationalize the findings, paradoxes remain. Why, for example, should those with fewest S responses be those who 'always have something to gripe about'? Why should those who 'make people feel comfortable' not be distinguishable on the basis of the major Rorschach signs from those who 'make people uncomfortable'?

Some of the relationships are large enough to suggest correspondence between Rorschach signs and social traits. More thorough studies using the reputation technique in well-knit groups might yield considerable information on the relation between overt behavior and personality structure.

SUMMARY AND DISCUSSION

The Group Rorschach records of two hundred students in the College of the University of Chicago were compared with grades earned, ACE scores, and criteria of social adjustment. The results were as follows:

1) Munroe check-list scores of first-year students correlated .17 with average grade on highly reliable and impartial comprehensive examinations. A subjective rating of adjustment as shown in Rorschach performance correlated .25 with grades. Neither of these correlations is high enough to enhance a multiple correlation at Chicago. Correlations for students in the second- to fourth-year classes were close to zero. A breakdown showing the interaction of Rorschach and ACE scores suggests that good adjustment may in part compensate for low academic aptitude, but this finding is based on too few cases to be accepted with confidence.

2) No statistically significant relationship could be found

between underachievement (grade average with ACE held constant) and any single Rorschach indicator.

3) Few single Rorschach indicators showed any degree of correspondence with total ACE score. None of the relationships was statistically significant, although some of them are in agreement with findings of previous studies.

4) These data do not confirm Munroe's finding of different Rorschach patterns for students with L higher than Q and those with Q higher than L on the ACE. No significant difference could be found.

5) Heads of dormitory units rated students on emotional adjustment. This rating correlated .31 with Rorschach adjustment rating. This is a significant departure from chance.

6) Sociometric descriptions of adjustment were obtained by having residents in the dormitory fill out a reputation questionnaire. The total rating of adjustment from this report correlated about .20 with Rorschach adjustment rating. This correlation is statistically significant.

7) A large number of specific hypotheses were tested, based on the presumed correspondence of single Rorschach check-list entries with traits described in the reputation questionnaire. The hypotheses based on Rorschach theory were confirmed slightly more often than they were positively contradicted, but a large number of expected relationships failed to materialize. The net confirmation of Rorschach hypotheses was not significantly better than chance expectation.

There can be little difference of opinion about the negative character of these findings. In the sample studied, the Group Rorschach, objectively scored, failed to predict scholastic success and gave only small correlations with criteria of social and emotional adjustment. Use of the Inspection Technique with the Group Rorschach for purposes of guidance in the College at the University of Chicago is not warranted by these findings.

This study does not substantiate the claims made for objective treatment of the Group Rorschach. The only appreciable correlations found are those relating the Group Rorschach to dormitory ratings. And correlations of the magnitude of .30 with such criteria can be obtained even with conventional questionnaires like the Bernreuter. The only published evidence, gathered under sufficiently controlled conditions and treated with acceptable procedures, which favors the Group Rorschach as a screening and prediction device in college is that from Sarah Lawrence. The differences between Sarah Lawrence and the University of Chicago are so great that the contradiction between the two studies may be explained on many grounds. Perhaps 'personality' influences grades at Sarah Lawrence, but does not affect the comprehensives at Chicago. Perhaps Munroe has such superior insight that her scoring of the Rorschach is a good predictor of grades whereas the Chicago scorers were not able to assign such valid ratings (but this suggestion is incompatible with the view that the Inspection Technique is an objective procedure.) Perhaps the selection of the Chicago group, or the different curriculum, accounts for lower correlations there.

This study emphasizes the futility of quick and easy hypotheses as to the meaning of Rorschach performance. One cannot look at a simple ratio like W:M and infer that the student will or will not perform as well as his ACE scores promise. If such a sign has any meaning, it has meaning only when the interpreter takes into consideration the age, sex, and level of intellect of the subject. Hypotheses that may be true for the average patient, or the average college freshman, must not be applied to very superior students until they have been carefully validated on such a group. Even such plausible conjectures as that students with emotional disturbance will make poor use of their ability, are too simple to do justice to the true operation of personality structure. In one student, severe maladjustment may be a factor causing him to concentrate and earn high grades, whereas equally severe maladjustment of a different character may impede the work of another student. Clinical study of Rorschach protocols may conceivably reveal such interactions. Even if the test is not useful as a statistical predictor, it may help the psychologically-trained counselor to guide students.

Further research with the Group Rorschach is recommended, particularly since the literature is still devoid of controlled studies on adequate samples in typical college programs. It is urged that such studies inquire what elements or patterns in the Rorschach are associated with particular behaviors and criteria, rather than attempting blind prediction of marks or over-all ratings.

REFERENCES

1) L. J. Cronbach. "Statistical methods applied to Rorschach test scores: a review." *Psychol. Bulletin*, Vol. XLVI, September, 1949, pp. 393–429.

2) E. A. Frankle. *The relationship between quantified Group Rorschach scores and sociometric adjustment ratings of a freshman college group.* Unpublished master's dissertation. University of Chicago, 1948.

3) F. D. Montalto. "An application of the Group Rorschach technique to the problem of achievement in college." *Journal Clinical Psychol.*, Vol. II, 1946, pp. 254–260.

4) R. L. Munroe. "Prediction of the adjustment and academic performance of college students by a modification of the Rorschach method." *Applied Psychol. Monographs*, No. 7, 1945.

5) ———: "Rorschach findings on college students showing different constellations of subscores on the A.C.E." *Journal Consulting Psychol.*, Vol. x, November, 1946, pp. 301–316.

6) H. A. Shoemaker and J. H. Rohrer. "Relationship between success in the study of medicine and certain psychological and personal data." *Journal Association American Medical Colleges*, Vol. XXIII, 1948, pp. 1–12.

7) G. M. Thompson. "College grades and the Group Rorschach." *Journal Applied Psychol.*, Vol. XXXII, August, 1948, pp. 398–407.

8) F. L. Wells. "Clinical psychology in retrospect and prospect." *Journal Psychol.*, Vol. XXVII, January, 1949, pp. 125–142.

9) J. R. Wittenborn. "Certain Rorschach response categories and mental abilities." *Journal Applied Psychol.*, Vol. XXXIII, August, 1949, pp. 330–338.

A NOTE ON THE EVALUATION OF COLLEGE REMEDIAL READING COURSES

H. A. ROBINSON

Clinical Psychologist, Division of Student Mental Hygiene,
Department of University Health
Yale University

I

Any investigator embarking on a survey of the extensive professional literature in the field of group remediation, cannot fail to note a striking lack of adequate validation data, particularly at the college level. A number of reports of experimental course findings may be uncovered in the several journals, but the standards of comparison used when remedial work is evaluated are, more often than not, specious, ill-considered, or clearly superficial. This is peculiarly true in the field of remedial reading. A review of nearly one hundred studies, for example, uncovers less than a dozen references to the effect of reading programs upon scholastic improvement. Of these only one study using control groups reports apparently significant gains in terms of academic grades for reading classes.[5] Other investigators either report no significant improvement in academic standing as the result of remedial instruction, or without definitive findings take the hopeful and confident stand that reading instruction can improve academic work.

This startling—and at the same time rather depressing—state of affairs poses a basic question for educators and research workers interested in college remedial reading. It would appear, even at this late stage of the game, that the effectiveness of reading programs is still largely suspect. Why is it that reading experts have bypassed what is seemingly a primary criterion for the evaluation of remedial instruction? Academic performance is clearly the *sine qua non* for the validation of remedial courses, particularly in liberal arts curricula where by far the largest portion of the scholastic agenda comprises reading or related activities. And in the final analysis remedial instruction must necessarily stand or fall on the basis of this single criterion, however ingeniously alternative standards of comparison are defended.

In speculating about the reasons for this consistent defection the tentative explanations which come to mind, plausible though they be, do not disguise the fact that remedial teachers (or, more properly, those people who concern themselves with the development of techniques and programs) avoid meeting the issue squarely when they skirt the problem of relating results to the well-defined life situation in which they operate, or concern themselves with other less meaningful values. *It is certainly true*, for example, that a simple demonstration of increased academic effectiveness as a result of group remedial instruction is not easy to arrange; under certain circumstances it may be well-nigh impossible. Most remedial reading courses carry no academic credit, and are of relatively brief duration. It may be that instructors are hesitant in their own convictions that a classroom learning experience of less, generally, than twenty-five class hours can have an observable effect upon overall academic performance, and so beg the issue by emphasizing simple gains in reading speed, increase reported in outside or non-academic reading, and the like. This in fact is an ostrich-like behavior which has had an inevitably unwholesome effect upon the development of worth-while remedial techniques.

It is the usual experience in group reading instruction that marked and statistically significant gains in reading speed do occur, whatever methods or techniques are applied. With his professional equanimity thus assured, the reading instructor may be loath to adventure into more indicative areas where less gratifying results would undermine the status and acceptability of the program. In certain environments this may conceivably be a practical and immediate consideration; in a broad view it is obviously short-sighted and deceptive. The fact is that the improvement of reading comprehension and the development of basic organizational skills are the core of the challenge which is posed to the remedial specialist. These abilities are complex skills which are accepted as being more closely related to basic intellectual endowment than the ability to read rapidly; accordingly, they are more difficult to revamp. Yet improved they must be, and with a demonstration that is convincing and stripped of idle embellishment.

It seems reasonable to wonder whether remedial teachers at the college level have an appropriate frame of reference for their work. Is the heart of the reading problem actually being approached? Are inadequate instructional methods being masked with a superficial show of improvement which is always conveniently at hand? Are the rôles of mechanical instruments and techniques being overplayed? Many remedial devices have a positive value—though their virtues are neither cardinal nor unlimited—and they are almost invariably effective for sustaining interest and motivation. Yet one suspects that because of their sheer and convincing mechanical facility, reading devices in the hands of some instructors are invested with the large burden of responsibility for reading improvement, and become an end in themselves rather than a limited accessory or handmaiden in the basic problems of reading improvement.

As for currently available remedial texts at the college level, no particular professional acuity is required to penetrate the superficiality of the types of exercises and treatments which characterize most of these volumes. The shallowness and malfunctioning of college reading programs at large is reflected (and at the same time reinforced) by the patent over-concern for a simple increase in reading speed, by the plethora of timed selections for the measurement of speed with no more than a passing regard for well-grounded index of comprehension, by the relative absence of working materials to develop basic organizational skills, and by the fetish usually made of a 'high level of interest,' which has produced a spate of materials and workbooks on a level of difficulty utterly impertinent to the task at hand. The continued demand of college educators for sound remedial work is made clear by the interest and support given the corrective programs which do exist; the fallibility of the programs themselves is reflected unhappily in the potboiling welter of generally inconsequential writing which decorates the bookshelves of remedial specialists and instructors. Put differently, there is some doubt as to whether the contemporary mass of study aids, workbooks, and corrective pamphlets in this area is really related to the problem of improving reading skills at the college level.

As a collective result of these elements, the status of group remediation in college continues to be unhealthy. Remedial work is still peripheral; remedial measures (and remedial specialists) are still under some suspicion. Integration of reading programs with the curriculum is tenuous if attempted at all. Students with reading difficulties are taught how to improve their scores on standardized reading tests, but are assisted to a far lesser extent in their basic problem of academic adjustment. Faculty coöperation leaves much to be desired, although some recognition and tongue-in-cheek acceptance is given in many universities. And the reading programs themselves to varying degrees are lacking in vitality, force, pertinence, direction. Why? Basically the reason is that remedial programs have not conclusively demonstrated their worth. In the eyes of educators at large this situation not only casts a mantle of doubt over the field of reading and the educational psychologists who are working in it, but as a matter of course it also has a systematically bad effect upon the life forces of remedial research. For it is of the nature of things that patent sterility in any educational field attracts neither the most capable of research workers nor the most effective and interested of instructors.

In gauging the value of their class work, remedial reading instructors in the past two decades have gone through a kind of developmental scale which may be sharply criticized at every stage. The first type of evaluation involved the simple and informal subjective appraisal of result by either teacher or student, or both. However indefensible this procedure from the standpoint of research accuracy, there is a deal of undeniable value in the sense that student and teacher were reasonably united in purpose. In verbalizing subjective reactions to a common experience—participation in a remedial class—the parties involved were compelled to express a frame of reference. And if points of view differed, they emerged and were immediately recognized—a condition by no means characteristic of more 'refined' evaluations.

A second step in appraising the results of remedial work was the classical and overworked test method, wherein alternate equated forms of a standardized reading test were administered before and after the training period. Standardized and objective reading tests with equated forms exist in large quantities, and test-retest results show statistically significant gains with gratifying regularity. This approach is useful only to the extent that a reading index or test score is a valid measure of the complexity of reading skills which a college student needs for a successful academic adjustment—a highly dubious assumption.

Actually, this type of appraisal does no more than demonstrate that remedial techniques may effect appreciable gains in 'reading test performance.' Aside from the basic artificiality of the procedure, there is a real danger that the organization of reading course materials will be directed to a perverted end, the improvement of reading test scores. Not only is the desirability of a meaningful transfer of acquired skills to an academic situation lost sight of, but in addition an insidious concept may be implanted in the mind of the participating student, who tends to acquire a false system of values with respect to his own reading abilities and skills.

A third stage of evaluation sets a comparison between groups of individuals given remedial training and comparable unexposed groups, in terms of standard tests or other measures. The introduction of so-called control groups is an artful salaam to scientific method, but at bottom this process is as unreal as either of the foregoing. It must be noted, however, that a leveling and leavening influence here does take place, since it is a common experience that the 'control' groups, given no remedial training, often show considerable gains—though not as great as those of the remedial groups themselves. The simple passage of academic time (or, more properly, the accumulation of certain environmental skills) appears to account for some 'improvement' on the basis of reading test-retest results. And, of course, for the thoughtful and conscientious remedial specialist the real question remains: Do reading tests measure the sort of thing he is attempting to improve?

Apparently the last step in the appraisal of reading courses (and this touched upon with unbecoming lightness in the literature) is the comparison of reading groups with suitably controlled non-participants, on the basis of academic performance. In these studies, reading class members are generally matched with other students for reading skill (test scores) and a measure of intellectual capacity. A comparison is then drawn between the two groups, in terms either of over-all scholastic performance or a suitable derivative (average grades in courses requiring heavy reading assignments, for example). With exceedingly rare exceptions these studies do not result in unequivocal conclusions. Differences between groups occur, and most often in favor of the remedial classes, but they cannot be defended on statistical grounds. This fact of itself may in part account for the infrequency of this kind of observation.

Of the approaches which have been briefly noted here, the last appears to be the most defensible. The standard of comparison is meaningful and real both for instructor and student; goals of instruction are necessarily well-defined in terms having an intense and practical value for an academic life-situation—so obliging the instructor to discard irrelevancies in organizing his material; the pitfalls intrinsic to the use of tests are avoided. All this notwithstanding, there appears to be one all-important objection to this evaluative approach. The unsound assumption is made that requirements of experimental control are satisfied if initial reading performance and intellectual capacity are held constant. That is to say, the proposition is made that there are no other important qualities which may distinguish students in remedial courses from those who do not participate. Initial attitudes and motivations are entirely disregarded; a singular omission and one which becomes all the more striking when it is remembered that most remedial classes are conducted on a voluntary basis, or at best have only partial academic credit. When students choose freely to present themselves for remediation, or are allowed a degree of choice in selection, the function of personality factors cannot in reason be denied. With a certain logic one could theorize, for example, that under those conditions reading course registrants might be overweighted with highly motivated individuals—the undergraduate 'eager beavers'—whose eventual performance gain would be associated, at least in part, with the qualities and traits which caused the selection of the course in the first place.

II

Why do students present themselves for remedial help? To throw some light on this question an informal survey was undertaken in the spring of 1948 at the Department of Psychiatry and Mental Hygiene, Yale University, where a remedial reading program for undergraduates has been developed. The subjects comprised sixty freshmen in Yale College, a random sample of a larger group who had participated in reading classes during either their first or second semesters. The procedures were simple and straightforward, including the group administration of a brief questionnaire designed for the purpose, and a half-hour individual clinical interview. The stated goal of this exploratory survey was to determine as accurately as possible the motivational patterns causative to 'voluntary' participation in remedial classes. Results were to a degree equivocal, and since the findings are subsidiary to the main issue of this note, a complete analysis will not be presented. Nonetheless, simply by way of illustrating the apparent complexity of student attitude facing the remedial teacher, there follows a brief summation:

For each of the sixty men seen, it was possible (by an objective tabulation of questionnaire responses and an evaluation of interview material) to assign a main determinant for the initial attitude toward remedial instruction. By no means entirely justifiable, this procedure tended to oversimplify the expressions, since often a primary attitude was seen to be contaminated or qualified by lesser dispositions. However, a substantial reliability was obtained in so rating these data for the principal attitude factor. Independent judgments made by two clinicians of both questionnaire responses and transcribed interview material resulted in perfect agreement in the selection of appropriate categories, and a percentage agreement of .93 in assigning subjects to the chosen categories (disagreements being referred to a third rater and a consensus obtained). It should be remembered that these students had given original lip service to the voluntary and non-academic character of the remedial instruction. In a word, they had presented themselves as aware of certain reading deficiencies and desiring improvement. Accordingly, the attitudes recalled as important at that time might be considered substitutive, although in terms of motivation value—and concomitant pedagogic challenge—they have been adjudged of primary significance. The expressions of these subjects may be assigned to six general categories.

1) For twenty-five individuals (little more than forty per cent of the group), the main reason for seeking remedial instruction appeared to be a knowledge of a reading disability and a sincere desire to improve scholastic efficiency through remedial work. In none of these cases was motivation markedly contaminated by other issues or values; these men were freely attempting to improve their reading skills, uninvolved with social and emotional

pressures about this step, relatively unconcerned with the self-devaluating judgments implied by their decisions.

2) In the case of fourteen students the primary reason for participation was a coercion (or a strong suggestion) on the part of a dean, a student counselor, or a faculty member. Although their reading deficiencies were clearly demonstrable, these men were without exception initially resistive to the idea of joining a special class and did so only to satisfy an official injunction.

3) For eight students, the reading course appeared to represent their last chance to remedy a variety of language difficulties which had been unsuccessfully dealt with—in all cases through individual work—prior to coming to Yale. These subjects were handicapped by speech defects, or specific spelling disabilities, or deficiencies in English usage, and while aware of the specific instructional goals, hopefully looked upon the course as being vaguely related to their own problems of language.

4) Declaration of candidacy for the remedial classes was, for six students, admitted to be an overt and conscious avoidance of certain demands which the university environment had made upon them, and which—in each case because of emotional or personality problems—they felt unable to meet. In effect the remedial work was the lesser of evils, a substitute activity. Thus, one student had been able briefly to extricate himself from compulsory physical training; another had parried a parental pressure to take part in the activities of a student-sponsored social service organization; still another was attempting to ward off the stigma of a fraternity rejection by methodically entering all available campus activities, academic or otherwise.

5) In a somewhat different way than the foregoing, the participation of four students appeared to relate directly to their inadequate and immature adjustment to the social and emotional requirements of a university atmosphere. They felt no specific anxiety or need; rather, an undifferentiated and diffuse concern for themselves and their work. Not markedly deficient in reading skills, they were 'joiners,' uncomfortable and inept, interested not so much in prestige values as in being occupied, in this case in a situation which offered a possible academic reward.

6) Lastly, the groupings above were not suited to three men, whose decision to take the reading course was for each a function of a basic personality maladjustment, and who were eventually

seen at the Division of Psychiatry and Mental Hygiene as clinical cases.

If they do no more, these observations serve to illuminate the source of some instructional problems in remedial work which are usually given little consideration in the evaluation of course results. They must in some way be treated if a realistic appraisal of remedial instruction is to be made.

It is noteworthy that with this representative sample of volunteer remedial subjects, less than half was of an initial persuasion which would make for productive and efficient teaching. This note is intended to suggest that the methods most often used to evaluate college remedial work are indefensible to greater or lesser degrees, and that any reasonable appraisal must take into account the attitudes and motivations of the participating students. One possible approach to the problem of motivation control is herewith offered in a brief account of a study conducted at Yale University.

III

An appraisal in terms of academic performance was made of remedial classes in reading conducted during the fall semester of 1948. The control group comprised a number of students who originally presented themselves as candidates for participation and who were unable to do so through force of circumstance.

As is the usual practise, a brief battery of reading tests was offered to all entering freshmen at the beginning of the term. Results were reported to all men, and an announcement of the offering of a remedial course (to carry no academic credit) was sent to those individuals whose reading speed and comprehension scores fell below the 20th percentile. At the organization meeting of the course approximately one hundred sixty students appeared to fill out a preliminary questionnaire pertaining to instruction, course detail, and sectioning. In view of the instructional load it was necessary to limit sections and to arrange course meetings arbitrarily. Of the men who wished to register, sixty were at that time unable to take part because of conflicting academic schedules or previous athletic commitments. The motivation toward the remedial work of those individuals who presented themselves as candidates and were in effect rejected, is held to be at least roughly equivalent to the introductory motivation of actual participants. Thus, if possible differences of intellectual ability and reading skill may be controlled, these two groups offer an appropriate means of gauging the effect of remedial instruction on academic performance. The motivation of the non-participants is, of course, under some suspicion; seemingly these men might have undertaken to transfer to other section meetings of academic courses, or to sacrifice a part of their athletic activity, in order to gain admittance. That their interest in the remedial course was somewhat more than transient, however, is attested by the fact that in a later semester forty-three of these students (seventy-two per cent of the group) registered for the work.

An initial comparison of participants and non-participants in terms of two reading test results, a rough measure of intelligence, and a combined index of college aptitude and secondary achievement, shows no marked group differences. Table I presents the mean scores for a measure of reading speed (words read per minute), a measure of reading comprehension, a general academic prediction, and the Verbal Section of the College Board Scholastic Aptitude Test (SAT).* Contrived as an index of verbal comprehension, the latter instrument is comparable to an index of (verbal) intelligence, and is often used as such for purposes of *general classification.*

The academic prediction scores are forecasts of average grade performance, and are derived both from the C.E.E.B. battery results and a suitably weighted secondary school grade average. Since as entering freshmen these subjects have no initial college grade standing, these predictions—combining academic aptitude and previous achievement—are particularly suitable as a standard of comparison.

The t values for these four measures show no significant dis-

---

* Most of the well-known group reading tests have been abandoned at Yale, *since the larger part of the population tends to array itself near the* upper end of the scale, making practical discriminations difficult. The reading tests used in this instance were two long excerpts (3000 words) from basic history and psychology texts. Twenty questions of a multiple-choice type were devised for each selection, and these tests, scored both for speed and comprehension, produced more satisfactory distributions than published standardized instruments. The Scholastic Aptitude Test scores, as a part of the regular C.E.E.B. battery, were conveniently available for each subject, as were also the academic predictions.

TABLE I.—GROUP CHARACTERISTICS; FOUR MEASURES

	Remedial Participants (N = 102)		Non-Participants (N = 60)		t	p%
	M	σ	M	σ		
Reading Speed (Words per Minute)	230	55.1	238	48.1	.94	30–40
Reading Comprehension	13.6	3.0	13.2	3.1	.81	40–50
SAT	538	27	529	28	.46	60–70
Academic Prediction	71.3	6.3	72.5	7.0	1.17	20–30

tinctions between the two groups. Differences which do exist may be attributed either to numerically inadequate sampling or—as is more likely—to some real motivational disparity. The closest approach to a statistically defensible difference is observed with the academic predictions, where an index between the 20 and 30 per cent confidence levels appears. Since these predictions are in part based upon a long-term scholastic performance, they are more likely to reflect motivation forces than the other three measures, derived from relatively brief test situations. It is interesting to note that the non-participant control group surpasses the remedial class members in two of the four measures. The control group displays a somewhat more rapid speed of reading, together with a slightly lower level of comprehension, than the experimental group. The mean aptitude test score is a little higher for the reading participants, who are at the same time assigned a somewhat lower average academic prediction. The combined significance of these latter two observations is that of a better secondary school achievement on the part of the control group.

In each case the SAT mean is well above the general college average of 500 on this scale, and somewhat below the Yale average of 550. Accordingly, this study centers upon a population average or high average in academic aptitude (unlike the burden of reading experiments reported in the literature, which have dealt with below-average groups). Reading speed and comprehension scores for both groups approximate the 25th percentile level for Yale freshmen populations.

The experimental group, disjoined into three sections of approximately equal size, was given twenty-one classroom hours of remedial instruction. The training period extended over ten weeks, and was completed shortly before the end of the half-year semester. Scholastic records for this academic period were collected and comparisons drawn between the remedial group and the non-participants. The two groups are contrasted in three ways; in terms of over-all grade averages, performance in courses of an essentially verbal nature, and performance in the residuum of courses (largely scientific or quantitative). The so-called verbal courses were arbitrarily selected to include those areas of instruction in which reading skills seem basic to scholastic success. The results of language courses (elementary or first-year classes excluded), and grades in Economics, English, Classical Civilization, History and other social science courses were combined to a verbal average. The remaining grades—heavily laden with science course results but including all elementary language—were also averaged. These data, with the *t* values, are presented in Table II.

TABLE II.—A COMPARISON OF ACADEMIC PERFORMANCE: REMEDIAL STUDENTS AND CONTROLS

	Remedial Participants		Non-Participants		t	p%
	M	σ	M	σ		
Total grade Average	72.4	7.1	70.8	7.3	1.4	10–20
Verbal Courses	73.0	7.3	71.2	6.8	1.6	10–20
Residual Courses	71.9	6.7	70.7	7.6	1.1	20–30

By no means an unimpeachable proof of remedial effectiveness, the results are nevertheless encouraging in view of the initial prediction that the control group would to some extent surpass the remedial students academically. The scholastic differences shown are in favor of the remedial group and, in the case of verbal course averages, the distinction verges closely on the 10 per cent level of confidence. To this extent participation is meaningfully associated with a higher academic efficiency. The least significant difference between the two groups appears, as might be expected, with performance in nonverbal subjects.

Inasmuch as these groups were originally seen, with the measures used, to be no more than roughly equivalent, the academic gain shown by the remedial students might be disputed. Accordingly a review of scholastic performance was also made in terms of carefully matched pairs of subjects. This procedure makes for an unavoidable sacrifice of population, but yields a near-perfect equivalence between a control and an experimental group. As many remedial students as possible were paired with nonparticipants on the basis of decile scores for the two reading measures and SAT results, and a three-point interval for the academic predictions. (In the latter case this arbitrary matching criterion was considered necessary because of the initial prediction disparity between the groups.)

Application of these standards furnished thirty-one matched pairs of subjects. Conformation between control and experimental groups is satisfactory; the t values for all four measures are smaller than .25 (less than an 80 per cent confidence level).

An examination of semester academic averages for these pairs of students shows a mean difference, in favor of the remedial group, on the order of a 10 per cent significance ($t = 1.7$). Further scrutiny of paired differences in terms of verbal and quantitative course performance discloses very similar distinctions. These results appear to buttress the observations made with the larger populations. To the degree that the motivations of the two groups for remedial instruction are comparable, the average gain of the remedial students is at least suggestive of a real academic utility.

SUMMARY

1) A review of current remedial practises at the college level reveals consistent shortcomings with respect to the validation of techniques. Academic criteria have been generally disregarded

in the evaluation of remedial work, and other methods, facile but less relevant, employed to demonstrate supposed improvement. Motivational factors involved in the student selection of remedial course work have been completely neglected in appraisals found in the literature. It is suggested that the unreal approaches to validation taken by remedial specialists have contributed in large part to the still ambiguous and ill-defined professional status of the field as a whole.

2) Results of a combined questionnaire-and-interview survey of student attitude toward a non-credit and volunteer course in remedial reading are reported. Initial motivation for the work ranged widely. Less than half of the group investigated displayed original motivation which was considered instructionally advantageous; the majority of students participated either through coercion or as a means of avoiding certain environmental demands.

3) In an evaluation of a classroom remedial program by means of scholastic performance, an attempt was made to take into account the initial motivation of participating subjects. Following the training period the remedial students showed average academic gains approaching significance over the control group.

REFERENCES

1) C. Bird, and D. M. Bird. *Learning More by Effective Study*, Appleton-Century, 1945.

2) Frederick, Kitchen, and McElwee. *A Guide to College Study*, Appleton-Century, 1947.

3) J. R. Gerberich. "Five Years of Experience with a Remedial Reading Course for College Students," *Journal of Experimental Education*, 1934–36, 36–41.

4) W. S. Gray, Editor. "The Appraisal of Current Practices in Reading," *Supplementary Educational Monograph*, 1945, No. 61.

5) R. W. Kilby. "The Relation of a Remedial Reading Program to Scholastic Success in College," *Journal of Educational Psychology*, 1945, 36, 513–534.

6) Mathews, Larsen, and Butler. "Experimental Investigation of the Relation between Reading Training and Achievement in College Composition Classes," *Journal of Educational Research*, 1945, 38, 499–505.

TRUANCY AND CLASSROOM DISORDER AS SYMPTOMS OF PERSONALITY PROBLEMS[1]

FRANCES A. MULLEN[2]

Director, Bureau of Mentally Handicapped Children, Chicago Public Schools

It is axiomatic in educational philosophy that the child who resorts to truancy from school or to disorderly behavior when in school does so because he is unhappy and unsuccessful in school. The reasons why he is unhappy and unsuccessful may be sought in his physical condition and history, in his innate mental capacities, in his home and community environment, or in the curriculum, teaching practices, and environment of his school. Any factor from one of these areas may be sufficient to produce a problem symptom; more likely several factors from several of the areas will be found to contribute to a particular child's behavior pattern.

Granting these premises, the major questions to be raised by this paper are these: Are the children who turn to truancy as a solution for their problems different from those who choose classroom disorder? What physical, educational, personal, and social factors appear to be more closely associated with truancy and which with disorderly behavior? Do these symptoms have a different constellation of associated factors for the adolescent child than for the pre-adolescent child?

The data for this study were compiled from the records of individual examinations of 1628 children, selected from the total of 16,944 records of individual examinations made by psychologists and psychiatrists of the Bureau of Child Study of the Chicago Board of Education in the calendar years 1945 and 1946.[3] To understand certain biases in the present data it is necessary to know more about the larger sample from which the 1628 cases were drawn.

In 1945 the staff of the Bureau administered 8262 individual examinations (not counting diagnostic reading examinations), and in 1946, 8679.[1] The subjects ranged from nursery school to college, with intelligence quotients from below 20 to above 200. Because no child may be placed in an ungraded division, or excused from school for mental immaturity, without the recommendation of the Bureau of Child Study, a large number of cases are referred to the Bureau because the school personnel believe it likely that the child is below average in intelligence. However, with the expansion of the staff and services of the Bureau in the past twelve years, such cases no longer predominate. Children are referred for every variety of school problem, including guidance of extremely gifted children, and sometimes at the request of parents or social agencies when no school problem has been apparent. This changed emphasis in the work of the Bureau is reflected in the increase of the mean intelligence quotient of those tested from 78 in 1937 to 88 in 1946. Boys have consistently predominated over girls in the ratio of about seven to three. About one-fourth of all cases in 1945 and 1946 were non-white. The present sample of 1628 cases includes children from well over half of the three hundred seventy-nine elementary and high schools of the public school system. Examinations by a large proportion of the seventy staff members are included.

---

[1] Paper read at Midwestern Psychological Association convention, St. Paul, Minnesota, May 7, 1948.

[2] Formerly psychologist, Bureau of Child Study, Board of Education, Chicago.

[3] Munson, Grace. *Bureau of Child Study and the Chicago Plan of Adjustment Service.* Chicago: Board of Education, 1947, pp. 52. Contains a brief description of the work of the Bureau and the nature of the examinations conducted.

A Bureau of Child Study examination usually includes an individual mental test, educational achievement tests, physical tests and observations, interviews with parents, school personnel, social agencies, truant officers, physicians and clinics, and others. Cases are cleared through the Social Service Exchange to determine all agencies which have had contact with the family. After an examination has been completed, the recommendations made, and the report written, many of the facts concerning the educational, physical, social, and family history and condition of the child are coded according to a definite numerical system[2] by the psychologist. Then an IBM punched card is prepared for each child, giving the numerical results of mental and educational tests, and the coded data on other factors.

For the present study, the punched card records of all examinations made in 1945 and 1946 were sorted. The records of all children of ages nine to sixteen inclusive who were coded as having histories of truancy but were not conspicuously disorderly in class were placed in the 'truant' group, while the records of all children of the same ages who were considered notably disorderly but not truant formed the group labeled 'discipline problem in the classroom.' Each group was subdivided into a group aged nine to twelve inclusive hereafter called the 'pre-adolescent' group and a group of thirteen- to sixteen-year-olds, hereafter called the 'adolescent' group. Obviously this is an extremely loose use of the term 'adolescent,' employed here for economy of expression in referring to the two groups, as probably descriptive of a majority, though not all, within the group.

---

[1] *Statistical Report of Services of Bureau of Child Study*, 1937-1946, Mimeographed. pp. 63.

[2] *Revised Code for Child Study Record, January*, 1945, Mimeographed. pp. 12.

In the statistics which follow, these four groups: the pre-adolescent truants, pre-adolescent discipline problems, adolescent truants, and adolescent discipline problems, are compared with each other and with the whole range of child study cases for one year, 1946. This whole group is, of course, far from a representative or random sampling of the school population. It is a group of children referred to an agency because they presented some type of problem in educational or social guidance. The truants and behavior problems here reported also are not necessarily representative of all truants or all the disorderly pupils in the school system, but only of those of the specified ages who were selected for referral to this agency.

SEX, FAMILY ORIGIN, AND AGE RANGE OF EACH GROUP

In the total group of all cases studied by the Bureau of Child Study in 1946, boys predominate over girls in the ratio of seven to three (Table I). In the two pre-adolescent groups here selected, they predominate in the ratio of about eighty-nine to eleven, suggesting at this age level a lesser tendency for girls to present these behavior problems than to present such other problems as apparent mental or educational retardation, or other causes for referral to a clinic. However in the adolescent groups, a very different situation may be noted, especially with reference to the truants. In the thirteen to sixteen age group, almost forty per cent of the cases included are girls. This probably reflects an actual increase in the incidence of truancy among older girls, but also reflects an added awareness on the part of the school per-

TABLE I.—DESCRIPTIVE DATA CONCERNING THE FOUR GROUPS OF CHILDREN STUDIED

	Pre-adolescent		Adolescent		All Children *Examined* in 1946 N = 8070
	Truants N = 224	Discipline *Problems* in Classroom N = 511	Truants N = 092	Discipline *Problems* in Classroom N = 231	
Percentage of girls in the group.....	11.2	12.1	30.4	16.0	20.8
Percentage of non-whites in the group	37.1	28.8	30.4	19.9	23.1
Mean CA in years and months......	11-3	11-0	14-11	14-6	11-1
Grade expectancy corresponding to mean CA........	6.3	6.0	9.9	9.5	6.1
Mean MA in years and months*.....	10-2	10-1	13-8	13-3	†
Grade expectancy corresponding to mean MA.......	5.2	5.1	8.8	8.3	†
Mean actual grade placement.......	4.5	4.7	8.8	8.2	4.8
Mean grade score on reading comprehension test‡..	3.4	3.7	6.4	5.7	†

* Based on Revised Stanford Binet, Wechsler Bellevue, or other individual mental test. Data not available for eleven to twenty per cent of each group. For most of these, mental data from a previous individual examination would have been available in the child's folder.

† Data not available.

‡ Based on standardized reading tests administered by the psychologist. Such tests were not administered for from thirteen to thirty per cent of each group of cases, usually because recent tests had been given by the school or sometimes because the child had no measurable achievements. Inclusion of such cases would be more likely to depress than to raise the mean grade level.

sonnel of the serious social implications frequently associated with truancy on the part of girls of this age. An adolescent girl truant is perhaps more likely to be referred to the Bureau of Child Study than an adolescent boy truant.

The percentage of non-whites (Table I) is higher in the groups selected for study than in the total of all child study cases, with the exception of the group of adolescent discipline problems. In the other three groups, approximately thirty per cent, or more, of each group are non-white, as compared with twenty-three per cent of the total group.

Less than two per cent of the children of any of our groups were foreign-born, but a wide variety of ancestral backgrounds are indicated. Children of Jewish descent are seldom found in the truant groups of this study, but much more often in the discipline problem groups. Children of Irish descent are more frequently found in the truant group than in the disorderly group. The data on the other national stocks do not indicate significant differences in the type of behavior symptom developed.

The ages of the various groups have been only partially controlled by the method of selection used. Since truancy as a school problem increases in prevalence with age, while classroom disorder apparently decreases as a significant factor in our series of case studies, our pre-adolescent truants have a higher mean CA than do the pre-adolescent discipline problems (though both were limited strictly to pupils of ages nine to twelve, inclusive). The same is true for the two adolescent groups.

MENTAL DATA

The mean IQ's of the four groups (Table II) range from 90.3 for the pre-adolescent truants to 96.1 for the adolescent truants, with all four groups including cases from below 60 to above 150 IQ.

Because of the still existing, though lessened, tendency of school personnel to refer to the Bureau psychologist chiefly children suspected of having IQ's below 80 (while handling problems of brighter children without the psychologist's assistance) these means must not be interpreted as though they were means for all the truants or all the discipline problems of the system. All four of our problem groups have higher mean IQ's than the whole range of Child Study cases. Certainly the mean IQ of 97

TABLE II.—PERCENTAGE DISTRIBUTION OF INTELLIGENCE QUOTIENTS IN FOUR SELECTED GROUPS OF CHILDREN

Intelligence Quotient	Pre-adolescent		Adolescent		All Children Examined in 1940 N = 8205*
	Truants N = 224	Discipline Problems in Classroom N = 511	Truants N = 602	Discipline Problems in Classroom N = 231	
Below 50	.....	.2	.....	.....	2.1
50 to 59	.4	1.0	.2	2.2	3.8
60 to 69	4.0	4.0	5.1	7.4	9.4
70 to 79	15.2	12.9	10.6	7.4	17 2
80 to 89	23.7	15.5	18.0	18.6	17.9
90 to 99	26.3	17.8	18.6	22.1	16.1
100 to 109	9.8	14.1	17.1	10.4	10.8
110 to 119	5.4	8.2	12.2	9.1	6.2
120 to 129	2.7	2.7	4.5	6.1	3.0
130 to 139	.4	1.0	3.3	3.0	1.3
140 to 149	.....	.2	.8	.4	.7
150 or over	.4	.4	.2	.4	.5
Not tested†	11.0	20.5	9.5	13.0	11.0
	99.9	100.0	100.1	100.1	100.0
Mean IQ‡	91.1	92.7	97.0	94.0	87.8
S.D.	14.39	17.0	17.9	19.3	20.5

* Children given audiometric examinations only have been excluded.

† For most of these children, intelligence quotients from an earlier individual mental examination would be available in the case folder, but was not available on the punched card record.

‡ Calculations of means and standard deviations are based on the number of children for whom intelligence quotients were available.

found for our group of adolescent truants strongly suggests that a random sample of this classification of child for the whole school system would show a mean of 100 or higher. The large percentage of children falling in the 80 to 99 IQ ranges suggests the possible hypothesis that many of these children, because of their personality disturbances, appeared to their own teachers and principals as being likely candidates for an ungraded division (below 80 IQ), whereas the psychologist's examination revealed that a dull mentality was not the root of the problem.

When mental status is compared on the basis of mental age (Table I), it must be remembered that in spite of the narrow range of age selection, the truants had higher average chronological ages than the discipline problems, within both the pre-adolescent and adolescent series. The differences in mean mental ages for the two types of behavior symptom are probably not statistically significant.

TRUANCY VERSUS DISORDER IN THE PRE-ADOLESCENT YEARS

As we survey Table III, we may first list the characteristics in which the truants show more serious problems than the discipline cases. Children of our younger truant group are more apt to come from broken homes than are the discipline cases (incidence fifty-two per cent and thirty-one per cent, respectively). A parent or sibling is known to be delinquent for 8.5 per cent of the truants and only 1.5 per cent of the discipline cases. More of the truants have a step-parent, more are in foster homes, and more are from families on relief. Known instances of lying or stealing are reported for 11.6 per cent of the truants and only 6.5 per cent of the discipline problems. The truants have had more checkered school careers in that they have changed schools more frequently and have had more periods of interrupted attendance, though they have not repeated grades any more frequently.

Second, we may note the characteristics in which the discipline cases show more serious problems than the truants, still confining our attention to the pre-adolescent series. The classroom discipline cases appear much more likely to be described as nervous and hyperactive, as employing attention-getting devices, as showing aggressive, anti-social behavior, and as possessing poor work habits. These things are probably partly descriptions of what is meant by the classroom disorder. They are very infrequently marked for the children who have records of truancy, but are not definite discipline problems in the classroom. A larger per cent of the discipline problems than of the truants of this age level are described as having unfavorable health histories, suggesting that the 'pampered sickly child' may be found in the discipline group more frequently than in the truant group or in the whole range of child study cases. Also a larger per cent of the discipline problems are the eldest child in their own

TABLE III.—PERCENTAGE INCIDENCE OF CERTAIN FACTORS AS CITED IN THE CASE HISTORIES OF FOUR SELECTED GROUPS OF CHILDREN

	Pre-adolescent		Adolescent		All Children Examined in 1946 N = 8670
	Truants N = 224	Discipline Problems in Classroom N = 511	Truants N = 602	Discipline Problems in Classroom N = 231	
Educational Factors					
Repeated grade or grades	64.3	60.3	53.0	55.0	30.6
Varied school experience	67.6	37.0	40.6	38.5	24.1
Interrupted attendance	34.4	20.7	26.7	19.9	16.8
Unsuitable placement, now or earlier	7.6	14.1	5.1	10.4	10.4
Reading disability	26.8	25.4	9.8	20.3	11.3
Skipped grade or grades	1.8	2.5	8.8	0.0	2.6
After school activities	3.6	1.2	7.1	6.9	1.0
More than one year over-age for grade placements	45.6	48.1	35.5	42.4	30.3
More than one year under-age for grade placement	.4	.2	1.5	.4	.4
Physical Factors					
Defective teeth	48.7	41.3	43.4	35.5	37.3
Defective vision	29.5	27.8	34.9	29.4	20.9
Ear, nose, or throat condition	20.1	16.2	17.8	16.0	16.1
Speech disturbance	5.8	5.9	4.1	0.9	10.9
Defective hearing	4.5	3.1	6.9	5.6	6.4
Unfavorable health history	4.5	11.0	6.5	6.1	5.7
Poor nutrition or general health	3.6	5.3	8.0	10.0	4.8
Behavior and Personality Factors					
Poor work habits	14.3	52.5	25.1	51.9	10.6
Nail-biting or thumb-sucking	5.8	8.6	13.1	10.8	7.0

TABLE III (*Continued*)

	Pre-adolescent		Adolescent		All Children Examined in 1946 N = 8679
	Truants N = 224	Discipline Problems in Classroom N = 511	Truants N = 662	Discipline Problems in Classroom N = 231	
Lack of self-confidence	4.5	8.6	19.6	13.4	6.0
Nervousness, hyperactivity	2.7	18.6	7.6	19.5	5.6
Social immaturity	3.1	5.9	5.9	7.8	5.1
Aggressive, anti-social behavior	3.1	27.4	5.0	27.7	3.5
Attention-getting devices	4.0	16.9	3.6	19.0	2.4
Habits of lying or stealing	11.6	6.5	9.7	4.8	1.9
Withdrawn, unsocial nature	2.2	2.9	5.6	4.8	1.8
Temper tantrums	1.3	3.3	1.8	6.1	.7
Position among Siblings					
Youngest child	24.6	23.0	25.2	25.1	25.3
Eldest child	13.4	22.0	23.7	23.8	21.2
Only boy among girls, or reverse	8.5	16.5	15.1	20.8	14.9
Only child	12.5	15.7	10.6	12.1	10.6
Sibling rivalry	2.7	3.8	4.7	9.5	1.4
Family Factors					
Broken home	52.7	31.4	48.5	37.7	24.9
Foreign language used predominantly in home	13.4	14.2	11.5	11.7	12.7
Step-parent in *home, now or formerly*	15.2	11.5	16.3	13.0	7.2
Crowded home	1.8	10.5	10.0	7.4	7.2
Unwise parental direction	21.0	18.6	27.9	27.7	6.9
Foster or boarding home placement	9.4	5.4	6.9	6.1	4.4
Inadequate parent	11.2	8.4	12.7	14.7	3.9
Family on relief	6.7	4.2	6.8	4.3	2.7
Parent or sibling delinquent or criminal	8.5	1.5	6.2	2.6	1.2
Insanity in family	1.3	1.3	2.4	3.0	.6

family, or are an only child, or are the only boy among girls or the only girl among boys, than is true for the truants. However, the incidence of these 'position-in-family' factors in the discipline group is practically the same as in the whole range of child-study cases, whereas the incidence among truants is definitely smaller, suggesting the hypothesis that such types of family position do not so much contribute to the development of behavior disorders as they militate against development of truancy.

Third, we may list the characteristics of equal incidence in both groups. A wide variety of physical handicaps are reported for all these children. The two groups do not vary significantly from each other or from the whole range of child-study cases in the incidence of specific types of physical defect, except in the general item 'unfavorable health history,' already noted. As to *educational factors*, over sixty per cent of both groups have repeated one or more half-grades, and over forty per cent are at least a year retarded in present grade placement, as compared with the normal grade placement for their chronological ages. Considering only the pupils given achievement tests by the psychologist, the groups have median reading-comprehension grade-achievement scores of 3.2 and 3.4, respectively, and median arithmetic scores of 3.9 and 4.1 (not shown in the tables), though they are actually in median grade placement of 4.5 and 4.7, and have apparently median mental capacity for academic work at 5.2 and 4.8 grade levels. In other words, both groups tend to be working more than a year below their actual grade placement in reading and more than a half-year below in arithmetic. If their achievement is compared not with their actual grade placement, but with their mental capacity, the deviations are even more serious. As to social and family factors, about twenty per cent of both groups of children are reported to have unwise parental direction (about three times as large a percentage as for the whole range of child-study cases) and in about ten per cent of both groups the parents are described as 'inadequate,' meaning that there is evidence in the case history of drunkenness, drug addiction, feeble-mindedness, ignorance, etc. Actual known insanity in the family is reported rather infrequently, and equally for both groups. A foreign language is used in about fourteen per cent of the homes of both groups, but this is about the same as the percentage for the whole range of child-study cases.

ADOLESCENT SERIES

When the adolescent truants are compared with the adolescent behavior problems, very similar relationships are found. The truants more frequently come from broken homes, more frequently from families on relief, or from families in which a parent or sibling is known to be delinquent. They have had more interrupted attendance at school, they show twice as great an incidence of lying or stealing. Factors more frequently associated with truancy than with classroom disorder at the adolescent level, which did not appear as significant differences at the pre-adolescent level, include nail-biting, lack of self-confidence, and inadequate recreational provisions.

When we look for characteristics which appear more serious among the 'classroom disorder' cases than among the truants, the same factors of nervousness, hyperactivity, poor work habits, aggressive anti-social behavior, and temper tantrums, mentioned for the pre-adolescent years, are found. There is still at this age a slight tendency for the only child or the only boy among girls or only girl among boys to show more discipline problems than truancy. However, the facts of being an eldest child, or of having a step parent, or being in a foster home no longer show significant differences of incidence between the two groups. In the area of educational factors, at the adolescent level, the discipline cases show more over-ageness, more discrepancy between grade placement and age, more reading disability, and slightly lower IQ (means 96.1 and 93.6, respectively). In this respect it should be noted that grade retardation and achievement deviations are apparently not as serious for the adolescent as for the pre-adolescent groups.

COMPARISON OF TRUANT BOYS AND TRUANT GIRLS WITHIN THE ADOLESCENT GROUP

Because of the increased percentage of girls within the adolescent truant group as compared with the other three groups of our main study, the girls and boys within this adolescent truant group were studied separately (there being four hundred one boys and two hundred sixty-one girls), to see whether any of the characteristics previously noted for the adolescent truant group should be considered primarily in terms of a different sex composition. The outstanding result of this breakdown was the very slight difference in percentage of incidence of most of the factors, between these particular boys and girls. Consequently the complete tables are not presented here. Larger percentages of the boys than of the girls were reported as having withdrawn, unsocial natures (6.7 per cent to 3.8 per cent); more of them also showed aggressive, antisocial behavior (6.5 per cent to 3.7 per cent); and more had speech disturbances (5.2 per cent to 2.3 per cent). In the realm of family factors, larger percentages of the girls came from families on relief, from crowded homes, from foster homes, and from homes where a parent or sib was known to be delinquent, while smaller percentages of the girls than of the boys came from foreign-speaking homes. The first set of facts is consistent with the rather widely hypothecated concept that boys show a wider range of behavior symptoms (from withdrawal to agression) than do girls; the second, with the experience of school people that a more serious constellation of home and school and personality difficulties must develop to lead a girl into an overt departure from conventional behavior.

SUMMARY

The children of our series of cases who have histories of school truancy or of conspicuous class-room disorder have been shown to be children with a high incidence of other factors. They have many physical handicaps (vision, hearing, teeth, etc.) though perhaps not more so than children presenting other types of school problems, such as retardation, lack of ability etc., who have not developed conspicuous behavior deviations. They have, however, had more erratic school histories than these other children who have been referred to the Bureau of Child Study, they come from homes presenting many more social and family problems, and they show poorer personal adjustment. Both types of children suffer extreme educational retardation as to grade placement, and even more as to achievement in the basic skill subjects. To what extent this retardation is cause and to what extent it is effect of the behavior problems it is, of course, impossible to say from such data as the present.

In general, children who choose truancy rather than classroom disorder as a method of escape from the tensions of the school situations have more home problems: crowded homes, broken homes, families on relief, parents or siblings delinquent or criminal; step-parents in the home, or foster-home placement. Such factors are particularly apparent among adolescent girls who are truant.

In adolescence, grade retardation and reading disability appear more prevalent among the pupils who have become conduct problems in the classroom than among those who have become truant.

In our sample of cases, both types of behavior symptom are associated with more serious educational problems (grade retardation and achievement deviation) in the younger group, and with more serious home and social problems in the older group.

There is some evidence that at the younger ages the fact of being the eldest of several children, or the only boy among girls, or only girls among boys, or only child, militates against the development of truancy but not against behavior disorders in the classroom. This effect of position in the family appears to be much lessened if not lost by the time the child becomes adolescent.

FACTORS IN STATE CHARACTERISTICS RELATED TO AVERAGE A-12 V-12 TEST SCORES[1]

K. S. DAVENPORT and H. H. REMMERS[2]

Purdue University

The A-12 V-12 qualifying examination described here was the first one given for the purpose of selecting a quota for placement in the college training programs of the two services. It was given in April, 1943, to more than 300,000 young men, either in high schools or colleges, and was probably the most extensive single day's testing with a single test ever held in this country.

The test was composed of verbal, mathematical and scientific items, and the scores had a close relationship with scores on the Army General Classification Test. Consequently, the test could perhaps be considered as a combination of a group 'intelligence' test and a general educational achievement test.

It seemed that the state means could give a rough indication of some sort of educational level of the various states. In a previous paper[3] the writer reported high correlations between these state means and certain educational finance data, for example, +.80 with current per pupil expenditures.

As a result of these data it seemed worth while to obtain the relationships between this test and other sociological, economic, and educational data. Table I indicates the nature of the variables and the size of the correlations. The variables were taken from John Gunther's *Inside U.S.A.* and give some indication of the 'goodness of living' of a state. When the state means were correlated with other data, as shown in Table I, it was evident that certain other variables were either measuring nearly the same thing or were closely allied with the test means. Since there was only one measure of educational achievement, it was

---

[1] A version of this paper was presented at the APA Convention in Denver, September 10, 1949.

[2] The authors wish to acknowledge the courtesy of Dr. H. T. Chauncey in allowing them to use the A-12 V-12 data and Miss Lois Remmers and Mr. Paul Baker for assisting in processing the data presented here.

[3] Davenport, K. S., and Remmers, H. H. "Educational Achievement as Compared with Money Spent on Schools." *School and Society*, 1945, 61:333–334.

TABLE I.—TABLE OF INTERCORRELATIONS

	1	2	3	4	5	6	7	8	9	10	11	12	13
1		.91	.57	−.63	.12	.15	−.67	.63	−.81	.88	−.50	.75	.81
2	.91		.49	−.58	.02	.19	−.63	.62	−.77	.81	−.51	.82	.76
3	.57	.49		−.40	.14	.40	.73	.16	−.61	.65	.08	.37	.67
4	−.63	−.58	−.40		.00	−.24	.69	−.44	.70	−.67	.25	−.45	−.53
5	.12	.02	.14	.00		−.09	−.02	.04	−.02	.05	−.09	−.21	−.01
6	.15	.19	.40	−.24	−.09		−.65	.15	−.26	.15	.16	.31	.01
7	−.67	−.63	−.73	.69	−.02	−.65		−.42	.74	−.75	.20	−.51	−.70
8	.63	.62	.16	−.44	.04	.15	−.42		−.51	.57	−.62	.70	.67
9	−.81	−.77	−.61	.70	−.02	−.26	.74	−.51		−.80	.36	−.59	−.66
10	.88	.81	.65	−.67	.05	.15	−.75	.57	−.80		−.50	.65	.83
11	−.50	−.51	.08	.25	−.09	.16	.20	−.62	.36	−.50		−.51	−.53
12	.75	.82	.37	−.45	−.21	.31	−.51	.70	−.59	.65	−.51		.69
13	.81	.76	.67	−.53	−.01	.01	−.70	.67	−.66	.83	−.53	.69	

1. Per capita income (1945)
2. Value of school property (1942)
3. Auto registrations/1000 (1944)
4. Lynchings (1882–1944)/100,000
5. Persons killed in auto accidents/100,000 (1943)
6. Killed in World War I/1000
7. Negroes/1000 (1940)
8. Residents/100,000 in *Who's Who* (1943)
9. Rural homes without privies/1000 (1940)
10. Telephones/1000 (prob. 1940)
11. Per cent population with no library service (1941)
12. Foreign-born/1000 (1940)
13. M score, A-12 V-12 Qualifying Examination (1943)

possible to deduce that such achievement might be accounted for, in large part, by other variables. Obviously it was not necessary to have thirteen separate variables to explain the correlations in Table I. A factor analysis thus seemed to be a logical procedure.

Certain difficulties, however, arose in making a factor analysis. It can be seen that many of the correlations are negative, and some highly so. The typical factor analysis begins with a positive matrix; there was no reason, however, why the original matrix should not have certain variables reflected before any factors were extracted. Variables 4, 7, 9 and 11 were so reflected, thus giving an essentially positive matrix. Extraction of the factors was then carried out by the usual procedure, one additional problem being recognized. Since most of these data were population data, there was a problem in knowing the size of the residual which could be accounted for in terms of chance. The writers arbitrarily set a residual of .10 or smaller as being a

TABLE II.—ORIGINAL FACTOR LOADINGS*

Variable	Factor			
	1	2	3	4
1	.930	.146	.204	.047
2	.894	.201	.067	−.067
3	.651	−.450	.396	−.371
4	.700	−.167	−.056	.190
5	.045	−.000	.238	.269
6	.325	−.496	−.451	−.301
7	.832	−.488	−.178	−.124
8	.095	.366	−.271	.120
9	.852	−.133	.071	.125
10	.913	.049	.185	.046
11	.496	.501	−.105	.267
12	.776	.342	−.273	−.315
13	.857	.238	.190	−.191

* *Variables 4, 7, 9, 11 reflected from original matrix (Table I)*

chance residual; after extracting four factors the procedure was stopped since there were only two residuals of the size of .10—all others being less than .10.

Table II gives the original factor loadings, with the notation that the variables which were reflected in the original matrix (Table I) were kept so deliberately. Since these were logical reflections, one need merely reverse the meaning of the variables to make them clear. For example, variable 4, number of lynchings per hundred thousand, when reflected, could be considered lack of lynchings per hundred thousand, and variable 7 would be lack of Negroes per thousand. These reflections should be remembered during the subsequent discussion.

As Table II is scanned, it can be seen that factor 1 has by far the heaviest loadings. The other three factors are much less highly saturated and perhaps are of lesser importance than factor 1.

Attempts were made to rotate the factors to meaningful structure, but, again, with all factors being essentially bi-polar, the typical procedure of maximizing on one factor and minimizing the negative loadings on the other factors proved to be relatively meaningless. A rotation to extract a 'principal' factor likewise proved to be of little value. Consequently the loadings were given a simple rotation through the centroid, resulting in the factor loadings shown in Table III. It will be noted that very little change was made by rotation in any of the factors, the greatest change coming in factor 4. After rotation, of the total contribution to the variance (sum of factor loadings squared), factor 1 contributes 72 per cent; factor 2, 14 per cent; factor 3, 8 per cent; and factor 4, 6 per cent.

The name of the first factor was seemingly obvious, but the other factors proved to be much more difficult for extracting a general principle for naming based on their loadings. Factor 1 is almost surely a state economic factor. Extremely high loadings are shown for variables 1, 2, and 10, with these variables being used frequently in determining a state's economic ranking. Loadings of .930, .894, and .913, respectively, for these factors (1, 2 and 10) indicate that these three variables are almost wholly measuring one thing.

Factor 2, with high positive loadings on variable 11 (note that this variable has been reflected) and variables 8 and 12, and with high negative loadings on variables 3 and 7—7 being another reflected variable—seem to indicate that this is a rural-urban axis, with rural on the negative side and urban on the positive side. States that were high in possession of libraries, in foreign-born, and in residents in *Who's Who* are predominantly urban; while states that have high automobile registrations and a low proportion of Negroes are predominantly rural.

Table III.—Factor Loadings after Rotation*

Variable	Factor			
	1	2	3	4
1	.030	.136	.202	.080
2	.898	.194	.069	−.035
3	.659	−.443	.403	−.353
4	.692	−.170	−.065	.211
5	.036	−.099	.228	.270
6	.330	−.492	−.448	−.298
7	.831	−.492	−.180	−.101
8	.694	.354	−.273	.152
9	.846	−.145	.064	.152
10	.912	.039	.182	.077
11	.493	.547	−.109	.289
12	.789	.344	−.261	−.286
13	.865	.236	.197	−.160

* Variables 4, 7, 9, 11 reflected from original matrix (Table I)

The third factor cannot be definitely identified, but it seems likely that it is a geographical factor, that is, a deep-South versus non-South axis. For example, states which are high in automobile registrations (variable 3), with a positive .396 loading, are not in the South; while states with a low proportion of soldiers killed in World War I, with a negative .451 loading (variable 6), are in the South. This latter fact may be accounted for, probably, in standards of selection and of soldier capabilities. The other positive and negative loadings seem to support this tentative naming.

We have not yet named factor 4—we are unable to pick out what is left, for example, in variable 3, when income, urbanity, and geographical location have been accounted for. The same question arises when variables 11 and 5 are considered. This factor is undoubtedly important, for it has substantial loadings on ten of the thirteen variables, and has a 6 per cent contribution to the total.

As a final statistic, the multiple correlation between the criterion of the A-12 V-12 test means and variables 3, 6, 8 and 10 is .962, indicating that from these four variables one could predict with little error the standings of the individual states on the test means.

It should be recognized that this test was given in 1943 and that there may have been great changes in the complexion of the states since then. It seems to the writers, however, that there is considerable food for thought in the results of the foregoing analysis; income is generally recognized as being a desideratum, if not a necessity, but it has not been so frequently recognized *that it is of such major importance in educational achievement.* When it is considered that the factor loading on factor 1 for the test is .857, that more than 70 per cent of the contribution to this educational variable is that of income, then a state's income must be heavily weighted in discussion of state educational achievement—if the hypothesis that this test is a reasonably good measure of such achievement be accepted.

These data are all state data; they do not apply to individuals. Without much facetiousness, however, we interpret these results to mean that the probabilities of reaching a high educational achievement are much greater if one comes from a high income state which is highly urban, which is not in the South, and which has such advantages as library service available to most of its population, has a high proportion of foreign-born citizens, a large number of residents in *Who's Who*, and many telephones.

USE OF THE MINNESOTA MULTIPHASIC PERSONALITY INVENTORY IN SCREENING COLLEGE STUDENTS FOR COUNSELING PURPOSES

GLENN R. HAWKES

Utah State Agricultural College
Logan, Utah

It has become apparent to college workers that many students fall short of realizing their full capabilities because of lack of adjustment. To meet this condition, more and more colleges and universities are instituting and developing counseling services. Because of the increased cost of such service and the extended time involved, it has also become evident that any device which improves the efficiency of counseling is highly desirable. Since the signs of latent personality disturbances and, in many cases, even existing maladjustment, are often not revealed in overt behavior, devices which would aid counselors in selecting from a large population the individuals having problems of personal and social adjustment or who are likely to develop these problems would be very useful.

In this investigation, the Minnesota Multiphasic Personality Inventory (MMPI), a device which has been found valid in other situations,[1,2,3] is evaluated with the intent of discovering how useful this inventory is in determining maladjustment in college life. Specifically, this study determines to what extent scores on this inventory earned by students when they enter college will be predictive of maladjustment which may develop later in several aspects of college life.

In an effort to measure the ability of the MMPI to predict as stated above, it was administered to the entering freshmen at Utah State Agricultural College, Logan, Utah, for the school year 1947–48. This group of a little over one thousand one hundred comprises the population from which the subjects were selected.

In order to derive data which could be tested for statistical significance, the following general procedure was used to select subjects for study:

It was found that fifty-two male and eleven female subjects deviated two sigmas above the mean on the MMPI areas of Hypochondriasis, Depression, and Hysteria. These subjects were selected for an experimental group. Due to disqualifying validity scores, non-availability of subjects, withdrawal from school prior to the setting up of the problem, the group was reduced to twenty-nine males and eight females.

This experimental group was matched subject for subject with twenty-nine males and eight females whose scores fell within the normal range as defined by the authors of the MMPI. Criteria for equating were as follows: same age, sex, class rank in college, and equivalent scores on the United States Armed Forces Institute tests of Effective Usage of English and Readings in Natural Science. Tests of significance of the equating scores revealed that the groups did not differ significantly. Such differences that were found later, therefore, were attributed largely to differences related to areas of the MMPI.

The selected subjects were then examined for differences in the following areas: (a) Interest, range and pattern, as revealed by the Kuder Preference Record, (b) Grade-point ratio for the academic year 1947–48, (c) Referrals to the Deans of the College for discipline, low scholarship, or commendation, (d) Ratings by faculty advisors using a rating scale, (e) withdrawals from school during the course of any quarter of academic work, (f) a single interview in order to determine social participation, objective in college study, overt manifestations of maladjustment, facility of expression, manner, etc. These areas were selected for three reasons: to the investigator it appeared plausible that students with personality disturbances might be expected to reveal their maladjustments in these ways. In at least one area (range of interests), previous research had indicated such areas as were studied were related to general maladjustment.[4] Finally, data on adjustment in these areas were available to the investigator.

The following procedure was established for deriving the data: The Kuder Preference Record which was administered in the battery given prior to registration was used to determine the fitness of the stated college objective. The stated college objective was determined in a single short interview. In determining the fitness of the objective with measured interest, the present objective was compared with suggested occupations for specified interest patterns. The Kuder Preference Record

Manual with the suggested occupations was used as the criterion. Arbitrary numerical values were assigned based on degree of fitness. The results were tabulated and tested for significance. A critical ratio of 2.81, significant at the one per cent level, was found which indicates there is a statistically significant difference between the experimental and control group in the fitness of their selected objective as compared with interest patterns yielded by the Kuder Preference Record.

On the basis of Berdie's research,[4] one might postulate that the neurotic person's interest is relatively flat when plotted vertically as opposed to the number of peaks shown for well-adjusted individuals. The nine Kuder Preference Record percentile scales were plotted on a Kuder psychograph. They were then assigned a numerical value as they deviated from the mean. This value increased at the rate of 1 point for each 5 percentile points. Values were ruled as positive in negative or positive deviation, inasmuch as they indicate either positive or negative peaks. The type of peak was not thought to be significant. A total score for peakness and flatness was thus obtained. A critical ratio of .88 was yielded in comparing the experimental and control groups. This is not significant but indicates a tendency in the direction indicated.

At the time of the interview, both experimental and control subjects were questioned as to their present college objective, previous objective if changes had been made, and their plans for college objective made prior to entering school. The results of these questions were then tabulated and given a numerical value. It was hoped that such determination would indicate the stability of the interests of the subject. Arbitrary numerical values were used in ascending order commensurate with the stability shown. The tabulated results of this phase showed no differences.

Adjustment to college life and the various demands made upon the individual by successful college work are no doubt reflected in scholastic achievement. To substantiate this a check of grade-point ratio was made. A critical ratio of 1.77 was yielded. This fails to be significant at the five per cent level but may indicate a trend of lower scholastic achievement for the experimental group.

It was stated in the hypothesis of the problem that maladjusted students would be referred to the deans most often for discipline or low scholarship. This premise was tested by contact with the deans and no significant differences between the number referred in the two groups was obtained.

Inasmuch as the faculty advisors are most likely to have information about a student, they were contacted regarding the subjects. To facilitate this information gathering and in order to preserve as much objectivity as possible, a rating sheet was devised. This rating sheet is typical of inventories such as the Haggerty-Olsen-Wickman behavior scale. The raters were cautioned to preserve objectivity as far as possible. In determination of statistical significance a critical ratio of 1.46 was yielded. This fails to reach statistical significance.

In a check of the area of withdrawal from college no statistical significance between the two groups was found.

Another phase of the experiment was to conduct a short well-controlled interview with each of the subjects. An attempt was made to determine the extent of social participation of the subject in the college and community. To preserve objectivity a random order of interviewing was selected and the experimenter was not aware whether the subject was experimental or control. The subject was asked about the extent of participation and then numerical values commensurate with this participation were assigned. A critical ratio of 2.53 was yielded which is significant at the five per cent level. This indicates that groups of better adjusted individuals participate more in social activities.

In addition to vocational objective and stability of objective, and extent and nature of social participation, the subjects were rated on facility of expression, appearance, attitude, and overt signs of maladjustment. Comparison of the group in this factor yielded a critical ratio of 1.69 which shows no significance but indicates a trend in favor of the control group.

SUMMARY AND CONCLUSIONS

A tabulation of all results is given in Table I which has been prepared in order to summarize and integrate the data presented. In two of the six instances, a significant critical ratio was found. The first hypothesis, that the control group would select objectives more in keeping with their measured interests, was confirmed. The difference here was significant at the one per cent level. For the fifth hypothesis, that the control subjects participated more socially, a difference, significant at the five per cent level was found. There exists a real difference between the groups in these two areas. The importance of these two areas to college adjustment, where the differences are significant, is indicated by Finch and Nemzek[5] and by Hill[6] who concerned themselves with social adjustment because of the weight attached to this element of college life. It also seems to the writer that a properly selected vocational objective which is closely related to the student's interests will assist him in adjustment to the college situation.

TABLE I.—TABULATION OF RESULTS IN THE USE OF MMPI IN SCREENING COLLEGE STUDENTS FOR COUNSELING PURPOSES

Area Measured	Mean		Crit. Ratio	Level of Signif.	Direct. of Diff.
	Exp.	Cont.			
Fitness of college objective compared with Kuder Preference Record	3.10	3.72	2.81	.01	Cont.
Peakness and flatness of interests as revealed by Kuder Preference Record	53.41	55.41	.88	Not	Cont.
Grade-point ratio	1.25	1.74	1.77	Not	Cont.
Numerical values of rating scales	20.20	16.61	1.46	Not	Cont.
Numerical values of social participation	79.04	84.32	2.53	.05	Cont.
Numerical values of social impression	20.91	22.64	1.69	Not	Cont.

In the remaining four areas where differences were computed, it was found that the difference in each case failed to reach statistical significance. It should be noted, however, that the trend in all of the measures was in favor of the control group. Possibly, had other areas of study been selected, differences in quality of adjustment might have been revealed more sharply. Failure to find differences of statistical significance in certain areas does not necessarily invalidate the device; it may lead one to question as to whether these checks were appropriate.

It seems justifiable to conclude, then, that some differences between the groups exist and that this device will predict these differences in at least two important areas. This indicates the value of using the MMPI as a screening device for college counseling. Students who are revealed by the device as maladjusted might well be assigned to an advisor skilled in personal adjustment counseling. This would more adequately utilize the services offered and make a guidance program more effective because stress is laid where it is apparently needed. Our results also indicate that the use of the device would not assure that all students in need of adjustment counseling would be detected, and that some students would be 'identified' who are found later not to need such help.

REFERENCES

1) L. R. Harmon, and D. N. Wiener, "Use of the Minnesota Multiphasic Personality Inventory in vocational advisement." *J. of appl. Psychol.*, 1945, 29, 132–141.

2) H. O. Schmidt, "Test profiles as a diagnostic aid: Minnesota Multiphasic Personality Inventory." *J. of appl. Psychol.*, 1945, 29, 115–131.

3) D. Bower, "The relation between intelligence and Minnesota Multiphasic Personality Inventory scores." *J. of soc. Psychol.*, 1947, 25, 243–245.

4) R. F. Berdie, "Range of interest." *J. of appl. Psychol.*, 1945, 29, 268–281.

5) F. H. Finch and C. L. Nemzek, "The relation of the Bernreuter Personality Inventory to scholastic achievement and intelligence." *Sch. & Soc.*, 1932, 36, 594–596.

6) R. Hill, "An experimental study of social adjustment." *Amer. sociol. Rev.*, 1944, 9, 481–494.

BOOK REVIEWS

FLORENCE L. GOODENOUGH. *Mental Testing: Its History, Principles, and Applications.* New York: Rinehart and Co., 1949, pp. 609.

Those who have admired and profited from the works of Dr. Goodenough throughout her career will be delighted that she has produced this volume. In it she looks at four decades of mental testing, lays bare the issues of most importance, and evaluates the progress that has been made. Her comments acquire impact both from her wisdom and experience, and from her no-nonsense style. Because this volume tries to cover all her viewpoints, the contents are extremely varied. There are sections on all types of testing, including motor development, ratings of social behavior, and projective tests as well as tests of the Binet genre. Because she equates critical thinking with statistical thinking, there is a series of chapters on sampling and statistical methods through analysis of variance. Several concluding chapters, much less meaty, survey the use of tests in schools, clinical practice, industry, social work, etc.

Goodenough covers knotty problems of mental-test theory which have been adequately restated in no modern book. There is, for example, a careful analysis of the scaling procedures of Heinis, Woodrow and Arthur, van Wagenen, and others. She makes considerable use of a provocative distinction between tests as samples of behavior, and test performance as a sign of traits. Another asset is a very satisfying history of early developments in testing.

Comparatively little space is given to describing tests or to citing evidence. Instead, Goodenough presents a personalized evaluation of each area. Often these comments are on the devastating side; scarcely a chapter dealing with work after 1916 is without such phrases as 'statistical illiterates,' 'insufficient controls,' or 'blunders and false conclusions.' Her position is conservative, and so will please some readers more than others. While she is open-minded on projective tests, for example, she chooses to approach them through a detailed consideration of Binet's case studies of his daughters, and then is quite explicit that the moderns have lacked the insight and the technique to do as well as he in the same area. Some of her treatment is open to criticism. She recommends using vocational aptitude tests of the Minnesota variety in a way the reviewer considers insufficiently critical. Her unqualified stress on the desirability of a normal distribution of test scores disregards recent thinking about maximum discrimination. Her suggestion that reading readiness tests are to be used in deciding whom to admit to first-grade does not consider their use in planning a suitable school program for every six-year-old.

While every reader will find something to disagree with in so uncompromising a book, no psychoogist or educator can come away from the book without fresh insights. It is particularly valuable to have attention drawn to some of the underlying problems and assumptions which are ignored when testing becomes a business or a technology. Because of Goodenough's perspective and tough-mindedness, her book will make a distinct contribution to the education of the advanced student.

LEE J. CRONBACH

University of Illinois

ALFRED A. STRAUSS AND LAURA E. LEHTINEN. *Psychopathology and Education of the Brain-Injured Child.* New York: Grune & Stratton, 1947, pp. 206.

Children physically handicapped or sound who show intellectual or personality aberrations as a result of injury to the brain substance is what this book concerns itself with. Content, methods used and general approach toward working with brain-injured children described in this volume will be of distinct value to all people interested in the education of the brain-injured child. The volume has been written by the president and the educational director of the Cove Schools for Brain-Injured Children at Racine, Wisconsin. The two authors have had nearly twenty years of experience in studying and working with these children, and they have learned much from these experiences. The book itself is divided into two parts. Eight chapters are devoted to psychopathology which composes Part I. Part II is composed of four chapters on the education of the brain-injured child. Also there is an appendix which includes an unpublished essay on the brain-injured deaf child written by Mary Agnes Blair, references and an index.

In general, this is an extremely well-done job. The part on psychopathology is clear, concise, illustrative, intelligible and comprehensive. It includes chapters on historical review of the literature, consideration of the anatomy of the brain and its function in the relationship to its foundations for considerations of some problems of the brain-injured child. Topics considered include perception and perceptual disturbances, thinking disorders, behavior disorders, testing the brain-injured child, clinic and diagnosis of brain injury on children, and the concept of the exogeneous child versus the endogenous child. The part on education includes consideration of general principles, and chapters on teaching of arithmetic fundamentals to the brain-injured child, teaching reading, and teaching writing. Reports of the results of their own studies and the evaluation of the contribution of such people as Kurt Goldstein and Doll are all well done. It is the most compact and comprehensive volume on this topic which has come to the attention of the reviewer.

H. MELTZER

Psychological Service Center
St. Louis, Missouri

D. SNYGG AND A. W. COMBS. *Individual Behavior.* New York: Harper and Brothers, 1949, pp. 386.

This book presents "A New Frame of Reference for Psychology," phenomenological psychology. "All behavior . . . is determined by . . . the phenomenal field of the behaving organism," not by objective reality, but by the situation as the organism perceives it. The task of understanding behavior is that of understanding or reconstructing the behaver's perceptual field. The task of predicting behavior becomes one of predicting how the behaver will perceive a situation at a future time; i.e., predicting his phenomenal field. There are two ingredients in the phenomenal field: the field of objective reality and the behaver himself—the 'phenomenal self.' Thus there are two ways open to changing behavior—changing the objective field and changing the phenomenal self. This general idea is as old as Herbart, at least; and except for the fact that 'apperceptive mass' is much less dynamic, that term might be substituted for 'phenomenal self.' The general idea is consonant with William James's declaration that "whilst part of what we perceive comes through our senses from the object before us, another part (and it may be the larger part) always comes out of our heads."

While phenomenological psychology lies in the field of field psychology and is regarded by some to be a radical branch, this book draws less upon field data than I anticipated. It draws freely from the data of traditional psychology. Except for some of its proposed experimental methodology, there appears to be little in it that is incompatible with functionalism or with Woodworth's general standpoint. Carr's treatment of perception and development of the self seems about as dynamic as that found in this book and in many respects is quite similar. I think one would be hard put to find many examples in Woodworth of the neglect of the individual in behavior which is attributed to objective psychology. However one may view the "new frame of reference"—perhaps it will be better received by clinicians than by experimentalists—it is a very competent work. It is fresh and crisp in style and thought. The educational psychologists and, for that matter, classroom teachers should find it particularly valuable. The chapters entitled, "The Goals of Education" and "The Task of the Teacher," are outstanding. Other chapters that should make a strong appeal to this same class of readers are "The Way We See Oursevles," "The Phenomenal Self in Action," "People Under Threat," and "Social Structure and Action."

The main orientation of the book is clinical. A good share of its latter portion is devoted to diagnostic and therapeutic procedures.

J. B. Stroud

State University of Iowa

Jean Paul Sartre. *The Emotions. Outline of a Theory.* New York: Philosophical Library, 1948, pp. 97.

In this short volume John Paul Sartre, better known for existentialism than psychology, attempts to outline a theory of emotions that can be read and understood but not in terms of a background of psychology as it is known in the United States today. This is instead a phenomenological study. In this little volume are considered the contributions and general significance of James, Janet and the viewpoints of Kohler, Lewin and Dembo.

Dembo is one of the heroes of the book, in a sense, since he more than James or Janet represents an orientation that Sartre thinks he follows. Of Janet he says that he was too uncertain, divided as he was between a spontaneous finalism and a fundamental mechanism, whereas in Dembo and Lewin he finds the first draught for expounding the pure theory of emotion and behavior. Sartre himself advises that the pages of the book should be regarded as an experiment in phenomenological psychology. In his words he says "we shall try to place ourselves on grounds of signification and to treat emotion as a phenomena," and this he does with the framework of concepts and words that will have some familiarity to American psychologists, but relatively little significance as far as influencing ways of investigation or interpretation. Illustrative of the nature of words and manner in which Sartre uses them in the book may be of interest. Some things are expressed in a startling manner as if they were revelations which are fairly obvious facts. On the other hand, other ideas are expressed in a flow of words where the meaning gets lost even though it is supposed to have significance.

Psychologists, says he, differ in their theories but agree that they should start before anything else with facts. Psychologists who might be in agreement about the fact that they should not accept their own existence in an *a priori* manner would be surprised to find that they are supposed to discover it in any other way but by observing that man is a man because he has the characteristics of a man. The general orientation and the search for the kind of significance that Sartre is known for is indicated in the last two sentences of the book which read as follows: "That there are such and such emotions, and only these, manifests without any doubt the factitiousness of human existence. It is this factitiousness which makes necessary a regular recourse to the empirical; it is this which, in all likelihood, will prevent psychological regression and phenomenological progression from ever coming together." And this is a tragic situation which few psychologists will bewail. Brilliant, some literary people have called this volume. Interesting reading to some psychologists this book can be. The curiosity of the reviewer is in what made Sartre take time off to write such a book. H. MELTZER

Psychological Service Center
St. Louis, Missouri

T. W. Richards. *Modern Clinical Psychology.* New York: McGraw-Hill Book Co., 1946, pp. 331.

The rapidly growing field of clinical psychology can profit from a literature which is modernized as well as validated. Large scale production and good quality control do not necessarily go together in clinical psychology any more than they do in other endeavors. In *Modern Clinical Psychology* Dr. Richards attempts to do a very difficult thing; namely, to write a book designed for professional students at a level that can be understood by the average college junior who has had at least one year of general psychology. What he favors is a dynamic approach, and by this he means an emphasis on the viewpoints connected with such names in the field of personality explorations as Murray, Rorschach, Klopfer, Freud and Jung. The clinical psychologist as considered in this book is concerned with the behavior of a particular individual, with the emphasis on the practical phases of behavior; in which respect the clinician resembles the clergyman, the personnel manager or the physician more than the general psychologist. The ability to understand others is viewed by the author as an intuitive gift, and the function of clinical psychology is to help improve this gift.

The heart of the content are seven chapters on appraisal of personality—two on the appraisal of capacity, two on the appraisal of motivation, and three chapters on the appraisal of control. These chapters on appraisal are prefaced by four chapters that serve as an orientation to the consideration of the problems of appraisal. The chapters include consideration of psychology as a specialty, methods of psychological appraisal, approach to the patient and physical examination. The appraisal chapters are followed by two chapters, one on precipitation and disposition, and one on readjustment. The book also includes a fairly adequate appendix, bibliography, and a list of visual aids as well as an index.

In one sense the book deserves to be called modern. It is different from previous books; it does pay attention to some of the more recent tools for the study of personality that are getting a good deal of attention—projective techniques, the Rorschach, the T.A.T., and the Minnesota Multiphasic. The author says the book attempts to integrate his personal experience in what he conceives to be the specialty of clinical psychology. There is no denying that this is his purpose. What can be questioned is the extent to which practice and literature has registered with the author and practice has been fully understood. In spots it is uncritically modern. There is nothing that says that everything modern is necessarily good, and everything that has a history behind it and has survived is necessarily bad. Comprehensive this book is not. If wisdom from broad living and reading is to be expected in a book on clinical psychology, one will be disappointed here. It does have a good deal of information about tools and knowledge, that clinical psychology does at the present time apply, which can serve to inform untrained people. The Multiphasic is over-used and uncritically so. Case study materials used do not signify the kind of an experienced clinician. Instead, they give the impression that it is a man who has searched for illustrations, rather than one who has selected them from experience. There is only one reference to Dr. William Healy's work, and this is a mention to the *Healy Picture Completion Test.* Was Healy not modern enough? The consideration of introversion-extratension concept is not critical. Reading disability is practically a neglected problem, though homosexuality is treated at some length.

All in all, what can be said of this book is that it can serve as a valuable reference book for students of clinical psychology. But the author made too many attempts to satisfy too many conflicting demands in selection and writing of the book. As it is, it is not deep, profound or understanding enough for graduate students of clinical psychology on the one hand, and presupposes much too much content for a person who has had only one year of clinical psychology to be able to make intelligent use of it on the other. The book, then, from the point of view of the reviewer, definitely does not satisfy the purpose for which it was written, but can be used to advantage as a reference work.

H. Meltzer

Psychological Service Center
St. Louis, Missouri

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 41 March, 1950 Number 3

PHYSICAL MATURING AMONG BOYS AS RELATED TO BEHAVIOR

MARY COVER JONES and NANCY BAYLEY
University of California

The problems of adjustment which are usually attributed to the adolescent period center around the youth's need to develop heterosexual interests, to select a vocation, and, in general, to acquire the status of adulthood in the eyes of his peers and of his elders. The impetus for the attainment of independent and mature status is undoubtedly related to the adolescent's physical changes, but the process of growing up is so complex and so interwoven with cultural factors that we have not yet been able to demonstrate more than a rather general relationship between physical and psychological phases of development.

It is well known that children mature at different rates, and reach the period of pubescence at different chronological ages. Although the psychological accompaniments of these differences in maturing can be examined in terms of mass statistics,[19,20] this approach to the problem is often disappointing because of its tendency to obscure the intricacies of the growth pattern and the dynamics involved in the process of integration. Case reports of individual children have been somewhat more successful in their attempts to disclose the processes involved in the attainment of maturity,[11,13] but the accumulation of individual life histories is a slow way in which to arrive at useful generalities.

The present report deals with two groups of boys who fall at opposite ends of a normal sample distributed on the basis of one developmental characteristic (skeletal age). In an attempt to find differentiating behavior characteristics, statistical comparisons of the two groups have been made and illustrative case material has been assembled for individuals falling at each extreme. The method, while providing no touchstone, does enable us to consider group differences* without losing sight of the individual behavior patterns of members of the group.

There are several ways in which children's physical maturity status may be expressed. One of the most commonly used for girls, is the age of menarche. As a possibly comparable measure for boys, some investigators have used the age of appearance of pubic or of axillary hair.[16] Shuttleworth[17] classified children in the Harvard Growth Study according to age at maximum growth in height. Height and weight have also been used as an index of physical maturity, although maturity differences may be obscured by genetic differences in measurements of gross body dimensions. This difficulty is avoided by the use of skeletal age norms, from x-rays of the long bones of the hand and knee.[21,22] Skeletal age has the advantage of being a stable and reliably assessed indicator of physical maturity, closely related to other aspects of physical maturing,[8] and applicable at all ages from birth to young adulthood.

Several studies have dealt with the relationship between skeletal maturity and intelligence.[1,18,24] Comparisons with other psychological factors are relatively scarce. In one of the few investigations in this field, Stone and Barker,[20] using the age at menarche as a criterion, found greater maturity of interests among girls who were past the menarche, than among premenarcheal girls of the same age.

1. PHYSICAL CHARACTERISTICS OF THE EARLY- AND LATE-MATURING GROUPS

The selection of contrasting extreme groups for the present study was on the basis of physical maturity assessments by the Todd standards for hand and knee. The groups included sixteen boys who were most consistently accelerated and sixteen who were most consistently retarded during the four and a half years for which we had cumulative skeletal x-rays, beginning at an average age of fourteen years. The total distribution from which these extremes were truncated consisted of ninety cases, a normal classroom sample of boys in an urban public school system.[13]

Figure 1 presents the comparative maturity assessments of the two groups at two age levels; each dot or circle indicates the skeletal age for an individual case. Some relevant physical data are given in Figure 2, which compares the distribution of the two groups in height at three age levels, and also with regard to adult height as predicted by the Bayley method.[3,5] The age levels used are grade placements (instead of more restricted age intervals) to facilitate comparison with the behavioral data which are of greater significance when kept in their original contexts.

---

* The significance of differences is reported for a given age, based on t ratios (Table I); and also for the seriatim data, based on the binomial test of significance (sections 2 and 5).

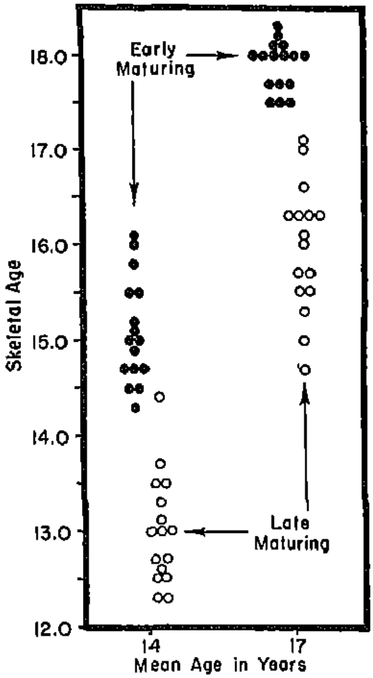


Fig. 1. Skeletal Maturity Assessments of Early- and Late-maturing Individuals, at Ages 14 and 17.

On the average, the physically accelerated and the physically retarded boys are seen to be of the same age, but are separated by about two years in skeletal age (the criterion variable). Although some overlapping can be noted (Fig. 2) in the height of individual children at each age, the means of the groups (indicated by arrows) are widely different. Even as early as eleven years (grade H5 and L6) all of the late-maturing are shorter than the mean for the early-maturing. At the mean age of fourteen years (grade H8 and L9) the distributions show an extreme separation; in the later years of adolescence the differences tend to decrease, and the predicted mature heights of the early- and late-maturing are very similar.

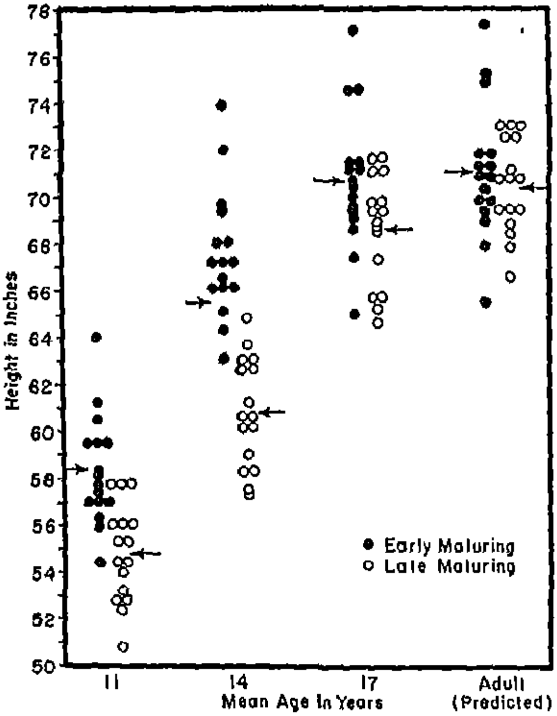


Fig. 2. Height Measurements of Early- and Late-maturing Individuals, at Four Ages.

There is also an obvious divergence (with no overlap) when the two groups are compared in terms of physical maturity ratings. This is seen when Greulich's five-point standards of maturity* are applied to photographs taken at fourteen years of age. The characteristic rating at this age is 3. The mean of sixteen early-maturing boys was 4.5 (close to the maximum), with no rating below 4. The mean of the sixteen late-maturers was 2.0, with only two ratings of 3.

---

* Based on ratings of pubic hair and external genitals.[9]

The boys who matured late were relatively very small from thirteen to fifteen years. In agreement with Bayley's[2] study of body build in relation to skeletal maturing, they were characteristically slender built and long legged at all ages. Furthermore, their strength tests show them to have been relatively weak at the ages when they were lagging in size, and their scores in the Espenschade tests of athletic ability[6] were in most instances below average. The early-maturing boys, on the other hand, were usually large, broad-built and strong, and tended to show good athletic skill throughout the period of our records. Their superiority in strength and physical skills was greatest at ages thirteen to fifteen,[12] when their early growth spurt accentuated their differences in size as compared with the slower-growing average and late-maturing boys. This is in agreement with a recent report by H. E. Jones, who presents a variety of data for groups of boys in the Adolescent Growth Study, considered in relation to static dynamometric strength. Strong boys were found to be relatively mature in skeletal age, weak boys were immature.[13]

2. SOCIAL BEHAVIOR IN BOYS' GROUPS

The psychological records examined in connection with the present study include both observational measures and reputation scores. We shall present first the ratings made independently by three staff members when the boys were in small groups (usually six) in a same-sex 'free play' situation. These will be referred to as ICW (Institute of Child Welfare) ratings.* The observations and ratings were concerned, in general, with social behavior and personal attributes which are important in social relationships.

The ratings have been converted into standard scores in which 50 represents the mean of the total group, with an SD of ten points. The direction and the degree of a child's deviation from the mean of his group are thus expressed in such a way that comparisons can readily be made between accelerated and retarded subgroups.

---

* Reference 14 contains a description of the rating scales and procedure.

Figure 3 and Figures 5–7 present cumulative standard score curves, from ages* twelve to seventeen, for a series of traits involving personal appearance, expressiveness, attention-seeking, and emotional patterns. As shown in Figure 3, the early-maturing are consistently rated as superior in physical attractiveness, with average scores which reach their highest value at age fifteen. In general, the group is about one SD above the mean total sample of boys. The late-maturing fall somewhat below the group mean, increasingly so from age twelve to age fifteen or sixteen. These differences in attractiveness of physique are complexly influenced by factors of size and of body build. Early maturing is not only associated with a more rapid growth in height, but also with mesomorphy.[10] The boys in this group tend to be 'well-built,' muscular, and athletic.† By contrast,

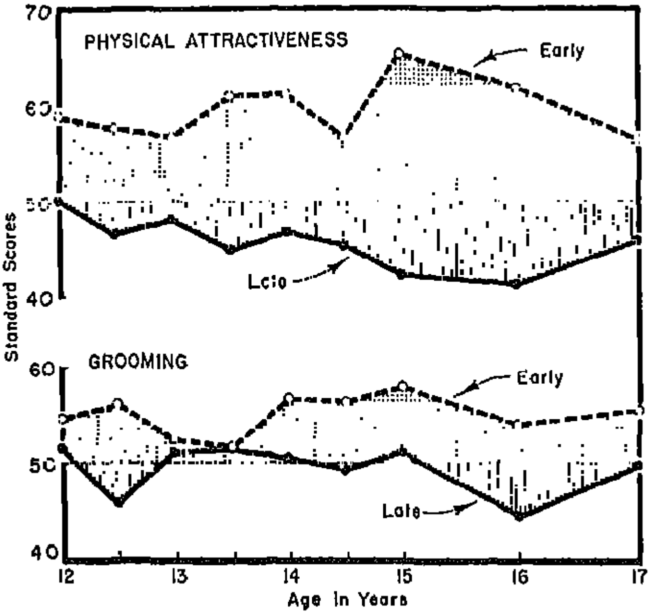


Fig. 3. Mean Standard Scores for Early- and Late-maturing groups, in Physical Appearance.

---

* The average ages are approximate, the groups at each age being based on a grade selection.

† By Sheldon's classification,[16] the average body-build formula was 2.8 – 4.4 – 3.5 for the early-maturing; 2.4 – 3.9 – 4.0, for the late-maturing.

the more slender, poorly-muscled build of the late-maturers was rated as relatively 'unattractive' by the adult observers.

When a comparison is made of the androgynic qualities of the builds of these two groups, the early-maturers are found to be on the average more 'masculine,' the late-maturers more 'childish' in their build.* As shown in the lower half of Figure 3, early-maturing, as might be expected, is also associated with a somewhat greater attention to the amenities of personal grooming. This is expressed in cleanliness, attention to hair and nails, and neatness of clothing.

The difference between the two groups in physical attractiveness is most marked at ages fifteen and sixteen. Figure 4 presents comparative distributions at age sixteen for a composite based on physique, grooming, and attractiveness of facial appearance. Each circle represents an individual case; the means are indicated by arrows. So marked is the separation of the early- and late-maturing that all but one of the former fall above the average of the latter; only two cases among the late-maturing are rated above the central tendency of our total group of boys.

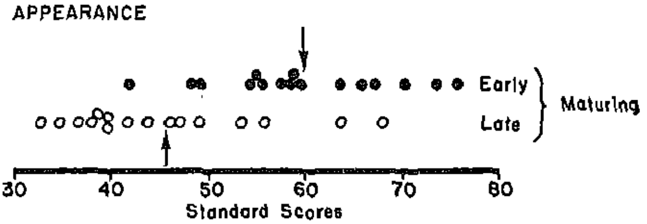


Fig. 4. Comparative Distributions of Early- and Late-maturing in 'Appearance' (means indicated by arrows).

Another group of traits which may have developmental significance are those related to expressiveness. Ratings of 'animation' and 'eagerness' are presented in Figure 5. In these characteristics the *early-maturing are close to the group average*, but the late-maturing are consistently above the average. Similar differences were found for other traits involving expressiveness; comparisons were made for behavior defined, at contrasting extremes, as talkative-silent; active-stationary; busy-idle; peppy-indifferent; and laughing-sober. In each of these the early-maturing boys were distributed similarly to the total sample of boys, the late-maturing were consistently on the 'expressive' side of the scale.

---

* *The androgyny ratings* were made *independently*, for another study.[4] 'Asexual' ratings (deficiency in masculine development) were much more common among the late-maturing, occurring in ten cases of this group as compared with only two cases in the early-maturing. Constitutional, and not *merely* maturational, factors are involved in these androgynic differences.

At least two factors are probably involved in determining this deviate position of the late-maturing. The first is a persistence of a childish activity pattern. A busy scurrying about and noisy interchange of shouts and comments is more characteristic of pre-adolescence than of later years; the adolescent often looks down upon such behavior as undignified, and adopts instead the rôle of a lounger, observing with tolerant superiority the childish antics of those younger than himself. A second factor is a reaction formation to inferiority. The 'active small boy' may be expressing through his activity not merely a survival of an immature culture pattern, but may also use this, as the only technique he knows, to hold the attention of others and to compensate for a physically less favored status.

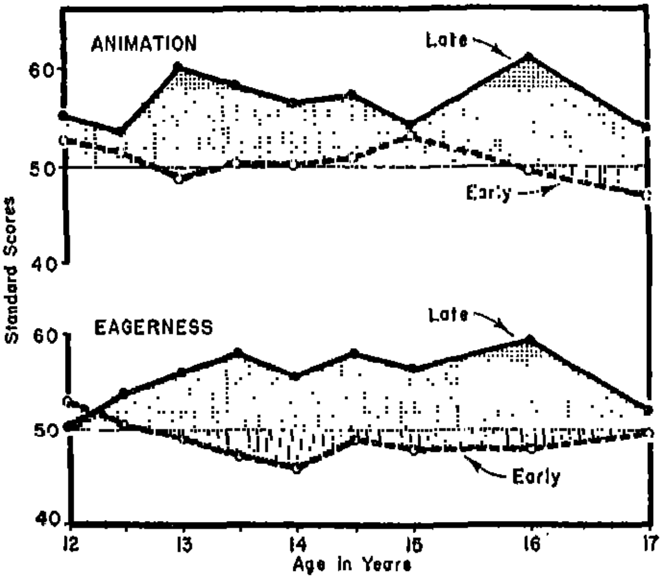


Fig. 5. Mean Standard Scores for Early- and Late-maturing groups, in Expressive Traits.

In this connection, it is instructive to consider the evidence concerning attention-seeking behavior, as presented in Figure 6. In the upper half of this figure, the late-maturing boys tend to vary around the average in the trait 'matter-of-factness.' Their lowest score is at age sixteen, when they fall on the 'show-off' side of the scale. In the lower half of the figure, their scores also vary around the average on the trait 'unaffectedness.' Again, they attain a low score at age sixteen; their expressiveness is judged to have a more affected quality at ages sixteen and seventeen than in the years immediately preceding. Although the differences are small, this would be consistent with an interpretation emphasizing a 'natural' or 'childish' expressiveness in the earlier years of adolescence, and a more compensatory attention-seeking expressiveness in the later years. In contrast, the early-maturing are at these ages judged to be relatively non-attention-seeking: unaffected and matter-of-fact.

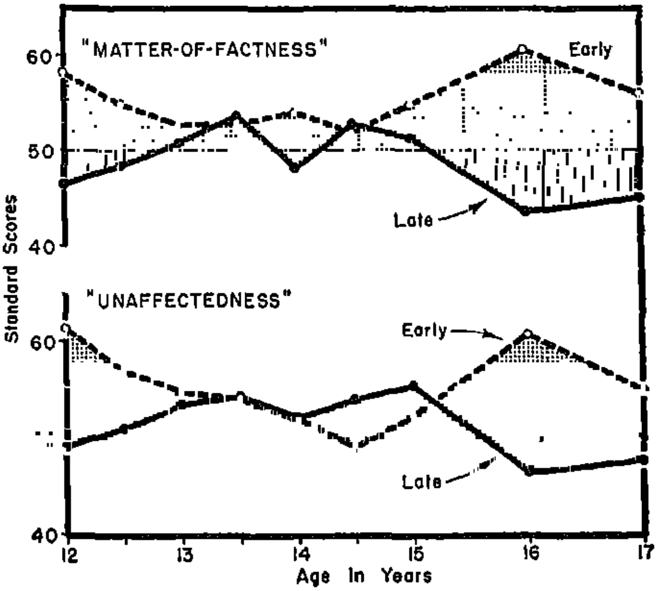


Fig. 6. *Mean Standard Scores of Early- and Late-maturing groups, in* Attention-seeking Behavior.

Also pertinent are the ratings for inhibition and relaxation in social situations (Figure 7). The late-maturing are relatively uninhibited, but they are also judged to be relatively tense. At age sixteen, the early-maturing are on the average approximately one SD above the group mean, in the direction of 'relaxation,' and the late-maturing are a similar distance below the mean, in the direction of 'tenseness.' The early-maturing are consistently well-adjusted in this trait, while the late maturing are in most semesters on the less well-adjusted side of the scale.

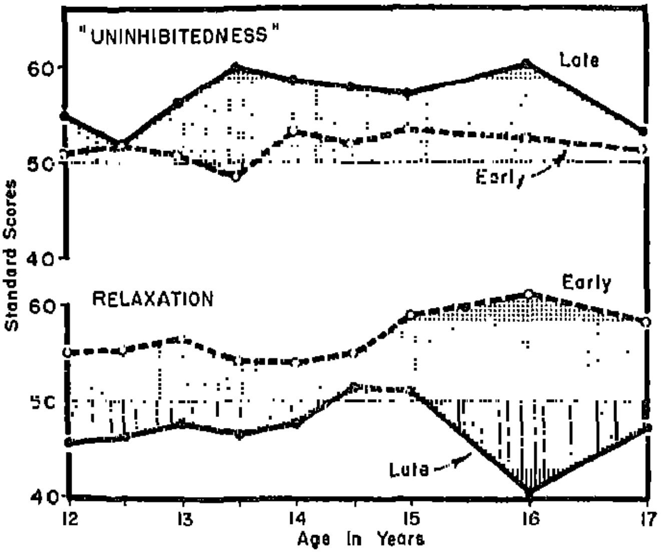


Fig. 7. Mean Standard Scores of Early- and Late-maturing groups, in Emotional Patterning.

Table I presents the significance of the differences between the early- and late-maturing at age sixteen, based on *t* ratios. For all but two traits on this list, differences are significant at the five per cent level or better. In these two traits (animation and 'uninhibitedness') significant differences are obtained when results are considered for all nine semesters in which ratings have been recorded, from ages twelve to seventeen, and analyzed in terms of the binomial test for consistency.* In attractiveness of physique, grooming, and relaxation (higher for the late-maturing) differences are in the same direction in each of nine semesters, with a significance level of .002. Differences of the same consistency were obtained for a number of other traits not included in Figures 3–7; these results show that the late-maturing are significantly more busy, more active, more 'peppy,' and more talkative. Differences at the two per cent level (in the same direction for eight of the nine semesters) were found for eagerness, social initiative, and sociability, in favor of the late-maturing. Differences at the seven per cent level (in the same direction for seven of the nine semesters) occurred in several additional traits, indicating a possibly greater tendency for the early-maturing to be good-natured, and for the late-maturing to be attention-seeking and to enjoy games.

---

* Based on the probability that the difference in a series of paired measures will occur in the same direction in a specified proportion of the instances.[7] Since the same subjects are involved, consistent differences are interpreted as independent of errors in measurement, but they are not necessarily independent of initial sampling errors.

TABLE I.—MEAN STANDARD SCORES FOR EARLY- AND LATE-MATURING

	Early	Late	Significance of Difference
Attractiveness of physique	60.6	45.0	.01
Grooming	54.6	49.8	.05
Animation	49.6	61.2	
Eagerness	47.9	59.3	.05
Uninhibitedness	52.5	60.2	
Matter-of-factness	60.5	43.6	.02
Unaffectedness	60.7	46.2	.05
Relaxation	61.1	40.6	.01

It may be noted that the two maturational groups show similar rather than different records in a number of traits of social importance. Thus, they present no marked nor consistent differences in observed popularity, leadership, prestige, poise, assurance, cheerfulness, or social effect on the group.*

In view of the relation of maturing to physical abilities, and of the high valuation placed upon athletic performance in the adolescent culture,[23] it is perhaps surprising that differences in maturing are not reflected in such traits as popularity, leadership, or prestige. Case reports[11,13] have made it clear that late-maturing is sometimes a primary source of social and personal maladjustment. On the average, however, the late-maturing boy succeeds in maintaining a fairly adequate status among his age-mates; very likely he is helped in this by his activity and other compensations, and it is also probable that some of the early-maturing are handicapped at times by the fact that they have outgrown their age group.

---

* As previously noted, the measures reported up to this point are based on behavior among small groups of boys. Somewhat greater prestige differences are apparent, especially in the later years, when measures refer to behavior in mixed groups, or to reputation scores based on votes from both boys and girls (considered in a later section).

3. DESCRIPTIVE DATA

In Figure 8, which presents unweighted composite values for various aspects of expressiveness* at age sixteen, it can be observed that six of the late-maturing boys fall quite outside the distribution of the early-maturing. A brief description of each of these individuals will provide a more concrete indication of a behavior pattern characteristic of nearly half of the physiologically retarded boys. This pattern of 'boyishness' is fairly well illustrated by the case of Tom:

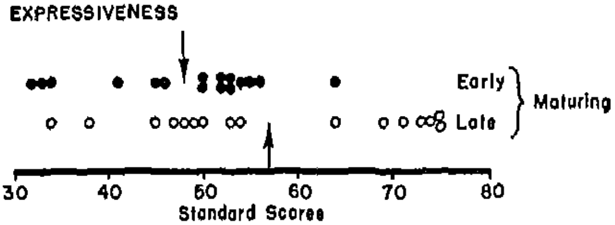


Fig. 8. Comparative Distributions of Early- and Late-maturing in Expressiveness (means indicated by arrows).

As a junior high school 'preadolescent' Tom was a chubby small boy, very rosy of cheek, sparkling-eyed, laughing and dimpled. His gaiety, good humor and dynamic activity were effective among his own equally boyish friends though most of his age-mates considered him to be quite immature in behavior.

Staff members made the following brief notations on the comment sheets recorded during free play situations (ages indicated in parentheses):

---

* The component variables are talkative—silent; active—stationary; peppy—indifferent; busy—idle; animated—stolid; eager—listless.

(13.4) In the nursery school yard: He seemed as delighted with the children of the nursery school as though someone had handed him a litter of puppies.

(13.4) At 'free' play with his own group: Tom played baseball with spirit if not with great skill. His activity was accompanied with continuous good-natured comments about the progress of the game and the 'boners' of his team-mates.

(13.9) Talking to another classmate: Tom Saylor chattered to Jim Cohn about his new job with as much enthusiasm as if he had found a gold mine. He is paid fifty cents a week to pass out handbills for a grocery store; this involves getting up at 5:30 every Friday morning. He was enthusiastic in urging Jim to "get in on it." The latter's indifference seemed to him like mild insanity.

(16.4) In senior high school: Still a small boy in comparison with most of his classmates, Tom's immaturity was obvious in many ways. His face was still cherubic, his lively and uninhibited activity more typical of younger boys. In the tenth grade an observer wrote: "He would seem to fit better in a junior high than a senior high school setting. Put him in a seventh or eighth grade classroom, and his appearance and behavior would be quite in keeping with that of the majority of his classmates."

Tom's clothes were misfits but he was as unconcerned about this as about his very dirty hands, although by this time most high school students were paying attention to clothes and cleanliness. His voice had not changed, as was very evident when he used it argumentatively in card games or jocularly in good humored razzing.

On the playground: Tom played pitching pennies with Ralph (an early-maturing boy) and seemed much more emotionally involved than the latter. He is often a conspicuous figure in childish scuffles; during a friendly tussle with several other boys, he was disturbed at finding that he had lost his penny. He was ready to cry and fight about this, the only coin he had.

During most of this period (junior and senior high school) Tom seemed not to suffer from introspective misgivings about his adequacy. By some magical formula, he apparently had caught and prolonged that legendary will-of-the-wisp, 'happy childhood.' Tom seemed oblivious to the tide of events which swept most of his classmates into an earlier puberty. But his euphoric and carefree attitudes were not shown to an equal degree by all of the other six late-maturing who were also classified as exceptionally 'expressive.' Among these with relatively similar patterns of conspicuous and active behavior were several boys who seemed more obviously to be compensating for a sense of physical inadequacy or for loss of a social prestige which they had enjoyed in earlier years:

Lonnie was a boy previously described by H. E. Jones[13] as one whose physical deficiencies in size and athletic prowess were a persistent source of tension and anxiety. His activity pattern was accompanied by excessive verbalizations and became more aggressive and compulsive as his status with the group declined and his home situation failed to support him during this difficult period.

Neal, another late-maturer, possessed a well-knit, athletic build, but was too small to play football or basketball on the team, as he had always been too small to compete with an older brother. His activity was chiefly limited to using athletic equipment; with other boys in a playground situation, he was inclined to be surly and on the defensive.

Bill and Teddy, two good friends from the same neighborhood, were active as nursery school children are active—they climbed, chased each other, hid and ran about. Often they were the ring leaders in mischievous hiding of the clothing or books belonging to other boys.

Milton was described as agile and quick, "climbed around like a monkey," " . . . his energy output seems to be twice that of an average healthy boy," he 'horsed around' in mixed group situations. In much of his group play he seemed to invite attack, to enjoy being teased and pummelled. At one time, however, he paid a mature classmate ten cents a week to protect him from some boys who were pestering him. An undisciplined child of somewhat elderly parents, he was disorganized, often rebellious, spent his energy in emotional outbursts as well as in random activity.

The six most 'expressive' late-maturing boys were conspicuous because of their activity in situations involving small groups of boys. They had chosen or were forced into what appeared, superficially at least, to be similar patterns of behavior. Yet the factors which determined the broad aspects of this behavior pattern were diverse and complex, altering in essential ways the details of the pattern from individual to individual.

Although expressiveness as exemplified by talkativeness, activity, eagerness and 'pep' is a form of behavior which many late-maturing boys easily assume or toward which they are predisposed, examples can be found of a quite opposite tendency. Two of the late-maturing boys are to be found near the extreme low end of the scale in expressiveness. These boys are brothers—Glenn and Charles. They were extremely self-contained and quiet, finding within themselves and their family satisfactions in behavior which did not require an audience or group participation.

Both boys were rated as non-attention getting and matter-of-fact; they were unsocial, and judged to have little effect upon their group of contemporaries. When most youngsters were seeking social satisfactions outside the family, Glenn and Charles were still enjoying congenial recreational interests with their parents. Their summer vacation was a family camping trip which involved packing into lonely country where the fishing and hunting was good. Glenn's response to the question, "Does your mother go camping with you?" was, "Yes, she likes to get away from people, too."

While in some respects Glenn and Charles resembled the early-maturing boys, their total complex of traits would never have led an observer to judge them to be among the more grown-up. In addition to having the lowest scores among late-maturers on 'expressiveness,' these brothers also had the lowest scores for 'sociability.' It is rather to be expected that girls would be outside their world. It was not until after they entered college that they began to take any initiative about dating girls. The mother, concerned with their 'backwardness' in this respect, tried to arrange for visits from families with girls of an appropriate age. After one such evening she was taken to task by her boys in this fashion: "Don't ever try that again. We didn't know what to talk about and it was a terrible bore."

Glenn and Charles are two boys who were late in reaching physiological and skeletal maturity and similarly slow to take on the social behavior of their peers. But in their slowness they remained aloof, bolstered by solid but solitary interests rather than by the energetic outgoing patterns of activity more common among the late-maturing boys of the study.

It is obvious that a statistical average which includes the social behavior scores of boys as different as Tom and the brothers—Glenn and Charles—will provide little understanding of the varieties of adjustment patterns adopted by late-maturing boys.

In analyzing the rating scores of the early-maturers, we find no clear-cut subgroup, such as our half dozen of conspicuously active little-boy late-maturers. This may be because the more grown-up boys are not required to over-react to an inadequate physical status and therefore have not tended to develop any one typical pattern in relation to this factor.

4. CLUBHOUSE RECORDS

Ratings and observations made in mixed group situations provide additional material for comparison of these two extreme groups of boys. These records were made in a variety of social situations, primarily in a clubhouse[10] maintained by the Adolescent Study near the school playground. In this situation it was possible to observe and rate a youngster's 'interest in the opposite sex,' based on approved behavior, participation in dancing and other mixed social situations, and talking about the other sex. As might be expected, marked differences were observed in this trait: in each of six series of records, beginning at age 13.5, the early-maturing obtained average scores above the group mean, and the late-maturing were consistently below the group mean. In this as well as in attention-seeking and other expressive behavior, a tendency could be noted for the late-maturers to show their widest deviation from the early-maturers during the period when they were most different in physical respects; in senior high school the apparent differences diminished, but as will be seen in the subsequent discussion of reputation records, the development of a more mature pattern, in the late-maturers, does not necessarily imply a prompt change in their status in the peer culture.

The narrative records obtained in the clubhouse situation provide evidence of the numerous specific ways in which maturity differences are expressed. At a Saturday graduation party three of the early-maturers came to the party conspicuously late, a mark of sophistication not shown by any of the late-maturers. Three others had Saturday jobs, had been working all day and were eager to talk to adults about their work. Among the observer's comments about the early-maturers were such notations as 'he dances well'; 'seems to think of himself as an adult'; 'acts a bit condescending'; 'reserved, little energy output'; 'Ran sat in the corner and flirted with Myra'; 'cheek-to-cheek dancing'; 'gay and assured, tried to cut in.'

In contrast, these phrases described some of the boys who were of the same age but were late-maturing: 'the first time I've ever seen him with a girl'; 'he held his head tensely while dancing'; 'didn't dance at all'; 'acted extremely silly'; 'Ronnie admitted he had been to only three dances before'; 'Claude showed a beaming countenance at all times'; 'began to wiggle and giggle.'

5. REPUTATION WITH CLASSMATES

Another source of evidence concerning adolescent behavior and status is from the Reputation Test. For the same age range represented in Figures 3–7, data are available from a series of tests in which classmates were asked to write down the name of anyone in the class conforming to certain descriptions.[23] For example, "Here is someone who finds it hard to sit still in class," or at the other extreme, "Here is someone who can work quietly without moving around in his seat." Scores were obtained by determining the percentage of times a person was mentioned on a given trait description, and these measures were then transformed into standard scores in which 50 represented the 'indifference point' (indicating no mentions at either extreme of the trait). Reputation scores are less differentiating than ratings; they tend to identify outstanding individuals, but may fail to distribute the middle range of cases who receive few or no mentions from their classmates. As a result differences between early- and late-maturing in average reputation scores are less marked than in average ratings by adults. However, a number of traits show differences which occur in the same direction on six testings, and are significant by the binomial test.

The late-maturing are consistently more 'attention-getting,' more 'restless,' more 'assured in class,' more 'talkative,' less 'grown-up,' and less likely to have older friends. On five out of six tests they are more 'bossy,' and less 'good-looking.'

Less consistency is found for traits which have been established as especially important for adolescent prestige. On judgments of 'popular,' 'leader,' 'friendly,' 'daring,' 'active in games,' and 'humor about self,' the late-maturing stand relatively well until the middle period of junior high school, and then tend to drop to lower status.

TABLE II.—MEAN STANDARD SCORES FOR REPUTATION TRAITS

	Early-maturing	Late-maturing
Attention-getting	48.1	52.2
Restlessness	45.3	52.9
Talkativeness	47.9	53.0
Bossiness	47.1	52.6
Assurance in class	45.6	50.0
Popularity	54.0	50.7
Leadership	51.3	47.5
Humor (about self)	53.5	48.7
Having older friends	56.2	42.3
Good appearance	54.4	49.3

Table II presents a comparison of average standard scores for the two groups in the H 10 and L 11 grades. The differences are not statistically significant but they present a picture which is in general similar to that already found in the observations by adults: the late-maturing appear as assertive (in a small-boy extroverted way) but at this age are somewhat lower in prestige traits.

At the earlier ages the more active and energetic of those in the late-maturing group were not unsuccessful in winning social recognition. But the early-maturing were much more likely to get and maintain the kind of prestige accorded to athletes and officeholders. Two of the sixteen early-maturing boys became student body presidents, one was president of the boys' club (a position next in importance to that of student body president), several were elected to committee chairmanships, and four attained outstanding reputations as athletes. The sixteen late-maturing boys produced only one somewhat 'important' officeholder (a class vice-president), and one athlete.

6. CONCLUSIONS

A general picture emerges from the various ratings and characterizations of these two contrasting groups of boys. Those who are physically accelerated are usually accepted and treated by adults and other children as more mature. They appear to have relatively little need to strive for status. From their ranks come the outstanding student body leaders in senior high school. In contrast, the physically retarded boys exhibit many forms of relatively immature behavior: this may be in part because others tend to treat them as the little boys they appear to be. Furthermore, a fair proportion of these boys give evidence of needing to counteract their physical disadvantage in some way—usually by greater activity and striving for attention, although in some cases by withdrawal.

In interpreting the relationships reported above, it must be borne in mind that we have used somewhat imperfect tools, and that a more complete psychological study of each individual could be expected to reveal maturity differences in manifest traits and in behavior dynamics which are not clearly shown in the comparisons we have been able to make. The question may also be raised as to the adequacy of our criterion for selecting the early- and late-maturing.

It is doubtful if any single event in the maturing process for boys can be compared in psychological importance to the menarche in girls. Skeletal age, as a measure of maturing, is relatively satisfactory with regard to reliability and stability. It is applicable at all ages from birth to adulthood, and is therefore well adapted to longitudinal studies of the same individuals throughout their period of growth. But skeletal age is revealed only by x-ray. Individual differences in this variable are hidden, and hence have received no cultural value-assessments. Feelings of perplexity or inferiority may arise, in relation to early or late skeletal maturing, only in so far as skeletal age is related to other physical features (such as size, strength, primary or secondary sexual characteristics) or to psychological changes in adolescence.*

It is not surprising that those who are retarded in skeletal development should often be directly aware of other aspects of physical retardation, and it is not surprising that this should lead to anxiety. We have seen, however, that there are many complicating factors, which make it difficult to predict the course which any individual adjustment will take. Some of our late-maturing boys enjoyed a degree of personal security, and status in other areas, which helped to balance their temporary physical inadequacies. A boy like Lonny, whose support at home was withdrawn at the same time that his retardation in growth became important, had a more serious problem than Tom, whose general security was adequate. Some of the early-maturing boys with fine physiques, nevertheless had disturbing accompaniments of rapid growth (such as severe acne), which tended to offset other advantages.

Our findings give clear evidence of the effect of physical maturing upon behavior. Perhaps of greater importance, however, is the repeated demonstration of the multiplicity of factors, psychological and cultural as well as physical, which contribute to the formation of basic personality patterns.

BIBLIOGRAPHY

1) E. M. Abernethy. "Relationships between mental and physical growth." *Monog. Soc. Res. Child Develop.*, 1936, No. 7, pp. 80.

2) N. Bayley. "Size and body build of adolescents in relation to rate of skeletal maturing." *Child Develop.*, 1943, 14, 47–90.

3) ———. "Tables for predicting adult height from skeletal age and present height." *J. of Pediatrics*, 1946, 28, 49–64.

4) N. Bayley and L. M. Bayer. "The assessment of somatic androgyny." *Am. J. of Physical Anthropol.*, 1946, 4, 433–460.

5) L. M. Bayer and N. Bayley. "Directions for measures and radiographs used in predicting height." *Child Develop.*, 1947, 18, No. 3.

6) A. Espenschade. "Motor performance in adolescence." *Monog. Soc. Res. Child Develop.*, 1940, 5, No. 1, pp. 126.

---

* Although skeletal age is highly correlated with other (more external) aspects of physical maturing, we must expect some regression toward the mean in such factors as growth of pubic hair, or other sexual characteristics, among boys selected at the extremes in skeletal maturing.

7) R. A. Fisher. *Statistical methods for research workers.* 10th Ed., 1946. Edinburgh: Oliver and Boyd, pp. 122.

8) C. D. Flory. "Osseous development in the hand as an index of skeletal development." *Monog. Soc. Res. Child Develop.*, 1936, 1, No. 3, pp. 141.

9) W. W. Greulich, R. J. Dorfman, H. R. Catchpole, C. I. Solomon and C. S. Culotta. "Somatic and endocrine studies of puberal and adolescent boys." *Monog. Soc. Res. Child Develop.*, 1942, 7, No. 3, pp. 85.

10) H. E. Jones. "Observational methods in the study of individual development." *J. Consult. Psychol.*, 1940, 4, 234–238.

11) ———. *Development in adolescence.* New York: D. Appleton-Century Co., 1943, pp. xvii + 166.

12) ———. "Skeletal maturing as related to strength." *Child Develop.*, 1946, 17, 173–185.

13) ———. *Motor performance and growth.* Berkeley: Univ. of Calif. Press, 1949, pp. x + 181.

14) F. B. Newman. "The adolescent in social groups." *Applied Psychol. Monog.*, 1946, No. 9, pp. 94.

15) H. G. Richey. "The relation of accelerated, normal and retarded puberty to the height and weight of school children." *Monog. Soc. Res. Child Develop.*, 1936, 1, No. 3, pp. 141.

16) W. H. Sheldon, S. S. Stevens and W. B. Tucker. *The varieties of human physique.* New York: Harpers, 1940, pp. xii + 347.

17) F. K. Shuttleworth. "The physical and mental growth of girls and boys age six to nineteen in relation to age at maximum growth." *Monog. Soc. Res. Child Develop.*, 1939, 4, No. 3, pp. 291.

18) K. Simmons. "The Brush Foundation study of child growth and development. II: physical growth and development." *Monog. Soc. Res. Child Develop.*, 1944, 9, pp. 87.

19) C. P. Stone and R. G. Barker. "Aspects of personality and intelligence in post menarcheal and premenarcheal girls of the same chronological ages." *J. Comparative Psychol.*, 23, No. 3, June, 1937, 439–455.

20) ———. "The attitudes and interests of premenarcheal and post menarcheal girls, *J. Genet. Psychol.*, 54, 1939, 27–71.

21) T. W. Todd. *Atlas of skeletal maturation (hand).* St. Louis: C. V. Mosby Co., 1937, pp. 303.

22) ———. "The roentgenographic appraisement of skeletal differentiation." *Child Develop.*, 1930, 1, 298–310.

23) C. M. Tryon. "Evaluation of adolescent personality by adolescents." *Monog. Soc. Res. Child Develop.*, 4, No. 4, pp. x + 83.

24) E. D. West. "Stage of ossification as a measure of growth and its relation to intelligence-test score." *Harvard Teachers Record*, 1936, 6, 162–68.

A STUDY OF RATIONALIZATION

ROBERT COBB MYERS

Educational Testing Service

Princeton, N. J.

I

Early in October, 1947, we administered an attitude-interest questionnaire to the members of the freshman class of an eastern women's college who had been admitted to the college just three weeks before. We will call these our postadmission respondents. Subsequently in 1947, and throughout the spring and summer of 1948, we administered the same questionnaire to applicants desiring to enter this college as freshmen in the fall of 1948. These girls answered the questionnaire at the same time as they were completing their usual application papers. The admissions office of the college selected the new freshman class from among these applicants without recourse to the questionnaire responses. After this selection had been made, we segregated the questionnaires of the successful applicants for analysis. We will call these our preadmission respondents.* Altogether, there were three hundred sixty-two preadmission respondents, and three hundred fifty-five post-admission respondents.

The questionnaire contained one hundred forty-five attitude-interest items in the following areas: motivations for going to college, study habits and teacher relations, interpersonal relations and interests. Answers were given on equal-appearing-interval scales of various lengths: twenty-nine items being provided with four possible responses, ninety items having five possible responses, and the remaining twenty-six items having seven possible responses. Examples of items having different response scales follow:

---

* These data were gathered as part of a larger study of nonintellectual factors in collegiate achievement being financed by the College Entrance Examination Board. The questionnaire was developed, pretested and administered under the general direction of H. S. Conrad. The insights of the following research staff members of the Educational Testing Service have been particularly helpful in analyses of the completed questionnaires: D. G. Schultz, W. B. Schrader, B. F. Green, Jr.

Four Response Scale

Item Example: How appealing would it be to you to join a group discussion on strategy for strengthening the United Nations?

Possible Responses:

1. Not at all interested.
2. Only slightly interested.
3. Moderately interested.
4. Warmly or quite interested.

First Five Response Scale

Item Example: How important was this reason in causing you to apply for admission to this college: The atmosphere is favorable to serious study?

Possible Responses:

1. Does not apply or is of no importance.
2. Of slight importance.
3. Of moderate importance.
4. Of considerable importance.
5. Of great importance.

Second Five Response Scale

Item Example: Do you feel that most teachers give you a "fair break"?

Possible Responses:

1. Yes.
2. Probably yes.
3. Uncertain.
4. Probably no.
5. No.

Seven Response Scale

Item Example: Do you feel discouraged after a session of studying?

Possible Responses:

1. Never.
2. Very seldom.
3. Rather seldom.
4. Sometimes.
5. Frequently.
6. Usually.
7. Always.

II

When the average responses of the preadmission cases were compared item by item with the average responses of the postadmission cases, we found that they differed significantly (CR of 2.00 or higher) in ninety, or sixty-two per cent, of the one hundred forty-five possible items. Table I shows the critical ratios which were found in this comparison.

TABLE I.—CRITICAL RATIOS OF QUESTIONNAIRE ITEM RESPONSE DIFFERENCES BETWEEN 362 PREADMISSION RESPONDENTS AND 355 POSTADMISSION RESPONDENTS

CR	Questionnaire Items		Cumulative Total	
	Number	Per Cent	Number	Per Cent
14.00–15.99	1	1	1	1
12.00–13.99	1	1	2	2
10.00–11.99	2	1	4	3
8.00– 9.99	4	3	8	6
6.00– 7.99	5	3	13	9
4.00– 5.99	28	19	41	28
2.00– 3.99	49	34	90	62
Below 2.00	55	38	145	100

In attempting to explain response differences of this magnitude between the Class of 1952 (preadmission respondents) and the Class of 1951 (postadmission respondents) one is faced with several options:

I) The differences reflect true attitudinal differences between the two groups which might be explained because:

(A) The groups might have come from different backgrounds and therefore were drawn from different populations.

(B) The preadmission students (Class of 1952) answered the questionnaire, on the average, several months after the post admission students (Class of 1951).

(C) At the time the preadmission students answered the questionnaire they were, on the average, about six months younger than the postadmission students had been.

(D) The preadmission students might have based their answers upon experience in high school, whereas the postadmission students might have based their replies upon experience in college (although they had had only three weeks of such experience at the time they filled in the questionnaire).

II) The differences do not reflect true attitudinal differences, but are simply an artifact which results because:

(A) The preadmission students answered the questionnaire in isolation, while the postadmission students answered in group situations.

(B) The preadmission students were under greater pressure to rationalize their responses since they had not yet achieved the goal of college entrance which they were seeking.

This list of possible plausible hypotheses illustrates once again the well-known difficulty that one can never statistically prove anything in social science; the best one can do is to find that the possibility of a given hypothesis being true has not been disproven. From our statistics alone we have certainly not disproven that any of the listed factors may not be at the root of the preadmission-postadmission response differences which we found. Although the magnitude of the differences may not by itself be able to resolve this dilemma, we can turn to the content of the items and judge which hypothesis seems to make the most sense, or be the most plausible, when compared to the quality of the differences.

We have carefully considered all the hypotheses in relation to the quality of the differences and have come to the conclusion that the hypothesis which seems to make the data as a whole hang together is the one which has been listed above as II(B). In another place* the writer has paraphrased Wallin to arrive at the general theorem that: "The interposition of a qualifying examination for a desired goal will ordinarily shift overtly expressed opinions in the direction deemed to be in conformance with achievement of the goal, i.e., in the direction of one's stereotype of the ideal goal achiever." When the pattern of the quality of the preadmission-postadmission differences within the ninety significant items is closely studied it appears clear that the preadmission respondents rationalized their responses in accordance with this theorem.

---

* R. C. Myers, "The Academic Overachiever: Stereotyped Aspects," *Journal of Experimental Education*, March, 1950.

We will now set down the qualitative differences found in the ninety significant items so that the reader himself may judge to what extent the hypothesis of greater preadmission rationalization seems to be the most plausible.

QUALITATIVE DIFFERENCES BETWEEN PREADMISSION AND POSTADMISSION RESPONSES

1. *Motivations for Going to College*

A) Preadmission respondents gave greater emphasis than postadmission respondents to these reasons:

Going to college trains one for community leadership, provides educational extra-curricular activities, gives training in sociability through living in dormitories or sorority houses, provides an opportunity of making a great many new friends and these friends are valuable help in postgraduation job hunting, gives opportunity to learn more about a great many interesting things, improves one's conversational ability.

Choice of this particular college was made because respondent had always planned to go here, the atmosphere is favorable to serious study, the college has high intellectual standards and is strong in one's chosen major field of study, one benefits from the desirable social background as well as a good education, many girls from the best families go to this kind of college and the type of girl one would wish to associate with socially goes here, most of one's friends go here, the college is not coeducational and it is known for its attractive girls, it gets one away from home, the type of man desired as a husband will prefer a girl who has gone here, and future children will be proud that their mother went to this college.*

B) Preadmission respondents gave less emphasis than postadmission respondents to these reasons:

College is a good place to meet a desirable husband, there is not much to do at home, and expectations of family and friends necessitate college attendance. The tuition and other expenses at this particular college are comparatively low.

---

* It should be noted here that although the college is an eastern liberal arts college, it is connected with a state university, draws its students from a broad cross-section of high-school graduates, and is by no means an 'exclusive eastern girls' school.'

2. *Teacher Relations and Study Habits*

A) Preadmission respondents presented greater agreement than postadmission respondents that they:

Had had teachers who generally explained things clearly, but that when something was not clear they asked the teacher about it, had teachers who gave them a "fair break," who seemed to care whether the students learned or not, and who stimulated the students to think about their courses, agreed with their teachers on what was considered important in their courses, had teachers who paid attention to them as individuals and who had made assignments clear, and that their teachers have appeared to like and enjoy the subjects that they teach.

Followed a study-plan or study-schedule, estimated prior to study periods about how much they were going to accomplish in the time available for each subject, and used odd moments (like time between classes) to review what they had learned, would be able to maintain an average of five hours per day of study in college if it were expected of them, believed that most of their school work had been rather closely related to their vocational goals, thought that examination grades really showed what one knew about a subject, believed that school marks or grades were generally useful and fair, but had confidence in their ability to pass an examination at the end of a course without studying for the exam, felt that the higher the grades a girl gets in college the more she will amount to after college and that grades are one of the most important aspects of school work, believed that the more one knows the more money one will eventually have.

B) Preadmission respondents presented less agreement than postadmission respondents that they:

Had found it difficult to understand what their teachers were 'driving at' in class, had teachers who failed to allow enough time to get assignments completed, found that their teachers had been unfair to them, and felt after an examination as if they had been tricked, accomplished more in courses where the teacher took a personal interest in them.

Lacked confidence in their ability to do satisfactory academic work, felt discouraged after a session of studying, and worried about their course work, read slowly, found it difficult to make up missed assignments and experiments or to get their daily assignments done on time, felt that much of what they had been learning in school was wasted effort and that many courses offered in college are not useful to a woman, found it was easy for someone to persuade them to do something else instead of studying and thought that anyone who studies most of the time is a 'grind,' were able to study some subjects considerably better than others and liked practical subjects better than theoretical subjects, found that they spent too much time studying one subject and then had to hurry through the others, studied much more efficiently under the pressure of immediate necessity and thought that 'cramming' was the best practical way to get good grades on a final examination.

3. *Interpersonal Relations and Interests*

A) Preadmission respondents rated these attributes of girl friends as being more important than the postadmission respondents rated them:

Mutual curricular and extra-curricular interests, good thinker (independent and original), has interesting ideas on current affairs and well-cultivated tastes in literature, interested in politics, mutual interests in similar business or professional career, and keeps appointments on time.

They also indicated greater interest than the postadmission group in these discussion topics:

Communism in theory and practice, peacetime uses of atomic energy in industry, some needed civic improvements, defects of the U. S. Bureau of Labor Statistics cost-of-living index, regulation of labor unions by law, current developments in U. S. foreign policy, strategy for strengthening the United Nations, the housing shortage, rival techniques of color photography in motion pictures, travel by airplane vs. railroad, and a recent popular novel.

B) Preadmission respondents rated these attributes of girl friends as being less important than postadmission respondents rated them:

Interested in dating, gay and carefree (doesn't take life too seriously), and lacks enthusiasm for the serious 'career-minded' kind of girl.

They also indicated less interest than the post admission group in these discussion topics:

How to pick a husband, peculiarities of one's school teachers, and recent news of happenings among one's friends or acquaintances.

After reviewing the quality of the preadmission-postadmission response differences, we see that the preadmission respondents preponderantly represented themselves as being more sober and industrious, better adjusted to the requirements of teachers, studying and examinations, having sounder motivations for college and presumably more acceptable reasons for wanting to attend this particular college, and more interested in having friends and discussions on the serious rather than the carefree side. That this is a result of greater rationalization on the part of the preadmission group, rather than representative of 'true' differences, we have little doubt.*

III

Not only did the preadmission respondents give responses which they appeared to rationalize as indicating 'better' attitudes on their part, but they also gave more extreme responses than did the postadmission group. This situation was anticipated, and stems from the hypothesis that overtly expressed opinions are apt to be more extreme when based upon rationalization than when based upon reality.† The ninety questionnaire items which had distinguished between the preadmission-postadmission responses with a CR of 2.00 or higher were, accordingly, tested against the following theorem: Considering that every questionnaire item was provided with a response continuum having a 'pro' terminus and a 'con' terminus, then whichever terminus, or pole, the average postadmission response falls nearer, the average preadmission response should fall still nearer (i.e., the direction of the preadmission response from the postadmission response should be further toward the terminus closest to the postadmission response). The result of this test showed that the conditions of the theorem had been met in the case of sixty-two (sixty-nine per cent) of the ninety significant questionnaire items. According to the Sign Test advanced by Dixon and Mood* we can consider our theorem of extremeness of response on the part of the preadmission respondents to have been upheld at the one per cent level of significance.

---

* One is reminded of the earlier report of Simpson that "there is a tendency for penitentiary inmates to rate themselves higher than college students in the possession of generally desirable personality traits." (R. M. Simpson, "Self-rating of Prisoners Compared with that of College Students," *Journal of Social Psychology*, 1933, 4, 464-467.) Simpson's prisoners wanted 'out'; our preadmission students wanted 'in.'

† The conventional interpretation of extremeness of response is that it indicates greater intensity of opinion. Among the most recent to present this viewpoint is Stouffer who postulates that: "People with the most extreme opinions, pro or con, also tend to hold these opinions with the greatest intensity." (Samuel Stouffer et al., *The American Soldier*, Vol. I, *Adjustment During Army Life*, Princeton, N. J.: Princeton Univ. Press, 1949, p. 43). This viewpoint may be sound so long as one is referring to covertly held opinions only. However, when one is discussing overtly expressed opinions, the hypothesis that greater extremeness of response indicates greater rationalization may be equally valid, depending upon the social situations under which the responses being compared were obtained.

A precise illustration of this situation would have involved determining experimentally what Stouffer would call a "region of relative indifference" on the continuum for the entire group of respondents, and then demonstrating that the responses of preadmission students tended to be further removed from this region than those of postadmission students. But, lacking this, we arbitrarily selected the midpoint of each response continuum as most likely being representative of relative indifference. Reference may be made to Table II for an illustration of our method. Here are presented 15 significant items, from a section with a five-place response continuum, in the same sequence as they appeared in the questionnaire. Using the mid-point answer ("Uncertain") as a reference point, one notes that in the case of 10 of these 15 items the direction and place of the arrow shows that the conditions of our theorem of extremeness of response on the part of the preadmission group have been met. In the case of 5 of these 15 items the conditions have not been met.

IV

An interesting by-product of this investigation has been the provocative finding that the magnitude of the Critical Ratio of the difference between the preadmission and postadmission responses seems to be directly related to the length of the response continuum which was provided for each item. Earlier in this paper we mentioned that, of the one hundred forty-five items in the questionnaire, twenty-nine had been provided with four possible equal-appearing-interval responses, ninety with five possible responses, and twenty-six with seven possible responses. When the data presented in Table I are broken down by these

---

* W. J. Dixon and A. M. Mood, "The Statistical Sign Test," *Journal of the American Statistical Association*, 1946, 41, 557–566.

TABLE II.--ILLUSTRATING THEOREM OF COMPARATIVE EXTREMENESS OF RESPONSE ON THE PART OF PREADMISSION RESPONDENTS

Item	CR
Do you feel you accomplish more in courses where the teacher takes a personal interest in you?	3.57
Do you like practical subjects more than theoretical subjects?	3.70
Are you able to study some subjects considerably better than other subjects?	3.39
Have your teachers seemed to care whether the students learned or not?	4.30
Do you read slowly?	5.62
Do you think that the more you know the more money you will eventually have?	2.42
If it were expected of you, would you be able to maintain an average of 5 hours per day of study in college?	7.92
In general, have your teachers stimulated the students to think about their courses?	5.46
Do you feel that the higher the grades a girl gets in college the more she will amount to after college?	9.50
Has most of your school work been rather closely related to your vocational goal?	3.14
Do you place importance on the idea that your college career will train you to participate actively in your community?	2.81
Do you think that anyone who studies most of the time is a "grind"?	2.63
Do you feel that most teachers give you a "fair break"?	5.76
Do you consider grades one of the most important aspects of school work?	2.25
Do you agree that "cramming" is the best practical way to get good grades on a final examination?	3.74

Responses

(Mean response of Postadmission students indicated by long line (|); that of Preadmission students by short line (ꞁ). Arrow indicates direction of difference.)

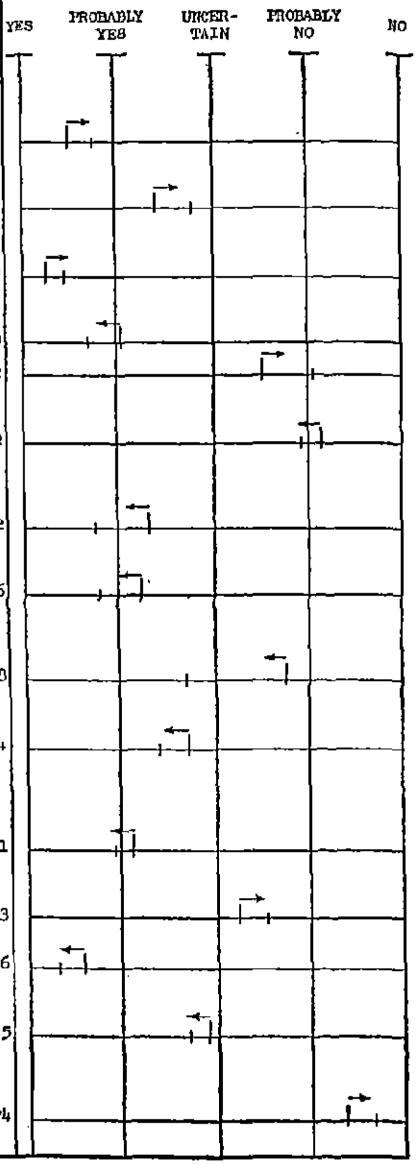


TABLE III.—CRITICAL RATIOS OF QUESTIONNAIRE ITEM RESPONSE DIFFERENCES BETWEEN 362 PREADMISSION RESPONDENTS AND 355 POSTADMISSION RESPONDENTS, BY LENGTH OF RESPONSE CONTINUUM

CR	Length of Continuum					
	4 Responses		5 Responses		7 Responses	
	Number of Items	Per Cent	Number of Items	Per Cent	Number of Items	Per Cent
14.00–15.99					1	4
12.00–13.99					1	4
10.00–11.99					2	8
8.00– 9.99			1	1	3	12
6.00– 7.99			3	3	2	8
4.00– 5.99	2	7	15	17	11	42
2.00– 3.99	12	41	32	36	5	19
Below 2.00	15	52	39	43	1	4
Total	29	100	90	100	26	100

three different response continua, we find the situation shown below in Table III.

The difference in distribution of CR's between items having a four-response continuum and those having a seven-response continuum is significant (CR = 4.32); also significant is the difference between the CR's of the five-response items and the seven-response items (CR = 3.55). The difference in distribution of CR's between the four-response and the five-response items would generally be considered not to be significant (CR = 1.71).

We realize, of course, that these rather striking differences may be a function of differences in item content between the various sections of the questionnaire.* However, there was a good deal of overlap in content between items having different lengths of response continua, and we suspect that the differences are, at least to some degree, a function of the length of the response continuum. To us, a reasonable explanation of this phenomenon would seem to be that there is a greater reluctance on the part of attitude-interest respondents to select a terminal (most extreme) response than some other optional response provided on a questionnaire. In the case of a seven-response item, the terminal responses can be avoided and still leave five responses from which to choose, thus allowing discrimination to that degree. However, in the case of a four-response item, if the terminal responses are avoided the respondent is left with only two options between which to choose, thus allowing but minimal opportunity for discrimination.

---

* This problem will not be resolved satisfactorily until we, or some other experimenters, have had an opportunity to try out items having different lengths of response continua, but exactly the same item content, on matched samples of respondents.

Using present Critical Ratio formulas, one might attempt to find the optimum length of the response continuum to yield the largest possible CR's for response differences between any two differentiated groups. Or, approaching the problem from another direction, one might set out to determine how the Critical Ratio formulas should be modified so that the CR will be held constant when nothing varies except the length of the response continuum.

SUMMARY

An attitude-interest questionnaire was especially designed for administration to entering female college freshmen. It has been administered to the members of two successive freshman classes at an eastern women's liberal arts college; in one case shortly after the students had been admitted to the college, and in the other case when the students were applying for admission. Of one hundred forty-five items in the questionnaire, the answers given to ninety were found to discriminate significantly (CR of 2.00 or higher) between the preadmission and postadmission groups. A qualitative analysis of these differences appeared to indicate that the preadmission respondents were more inclined to rationalize their responses than were the postadmission respondents. It was also found that the preadmission group gave more extreme responses, and this was shown to be significant at the one per cent level. From a subsidiary finding the question is raised as to whether or not differences found in magnitude of CR's may be a function of the different lengths of response continua provided for different section of items.

HOSTILITY, COMMUNICATION, AND INTERNATIONAL TENSION:
II. SOCIAL GROUP BACKGROUNDS

HARRY A. GRACE
University of Illinois

A. HYPOTHESES AND INTRODUCTION

This study is designed to find what relationships exist among expressions of hostility in three kinds of social situations: everyday, professional (teaching), and international situations. The hypotheses are that: (a) different categories of hostile responses exist in given social situations; (b) measures of hostility among the three different kinds of social situations are related; and (c) there are differences in social backgrounds between those groups scoring at the extremes in types of social situations and categories of hostility. The development of the instrument used in this study, a discussion of its reliability, and of the results which refer to hypotheses (a) and (b) are reported in an earlier paper.[5] This paper is concerned with the results of the comparisons of the social backgrounds of subjects scoring at the extremes on this inventory.

B. METHOD AND PROCEDURE

The instrument used in this study includes short verbal situations in three fields* of social behavior: (a) everyday, (b) professional (teaching), and (c) international. Everyday situations include stubbing one's toe, having a letter returned unmailed, etc. Professional situations concern problems of classroom discipline, student-teacher relations, etc. International situations center about relations between the United States and other countries, and include such things as the death of an American in China, the British tax on American films, etc.

After a preliminary test four categories† of hostile response were formed: autohostile, (expression of hostility against the self or the in-group), *laissez-faire* (allowing the action to continue without interference by the subject), verbal-heterohostile, (using verbal abuse against a target away from the self or in-group), and direct-heterohostile (directly attacking a target other than the self).

---

* Hereafter these areas of social behavior will be called fields.

† Hereafter these responses will be referred to as categories.

Thirty situations in each of the three fields with the four categories in random order after each situation were given to two hundred ten students at Teachers College, Columbia University in the spring of 1948. The average age of the group was thirty-one. The group was heterogeneous in background, having come from all parts of the country. The following social background data were solicited anonymously: sex, age, marital status, veteran status, parents' nation of birth, subject's nation of birth, parents' annual income,* subject's annual income,* parents' occupation,* subject's occupation,* and subject's political and religious preferences. (See Table I.)

C. RESULTS

After the item analysis there remained twenty-two situations in each field for final scoring, a total of sixty-six situations. The reliability, validity, and differences among the various fields, categories, and field-categories† are discussed in other papers.[4,5] Since the relationships between fields and categories are important for full understanding of the importance of social group backgrounds, the data are summarized below.

The three fields were not found to be significantly different from each other. The categories of hostility, however, differ very significantly from each other. The subgroup interaction, represented by the field-categories, also showed significant differences.

Social backgrounds of the high twenty per cent and the low twenty per cent of the subjects scoring in each of the field-categories were compared.‡ A difference of twenty-two percentage points is necessary for significance on the five per cent level, and twenty-eight percentage points' difference is necessary for significance on the one per cent level.[18] Eighteen percentage points' difference has been taken as approaching significance on

---

* These data could not be included since so many of these questions were unanswered by the subjects.

† The word field-category refers to the relationship between any field and any one of the four categories, e. g., everyday-autohostile, international-verbal-heterohostile, etc.

‡ Full table of these data may be found in [4].

Table I.—Summary Table of the Socioeconomic Groups Represented in this Study. Raw Numbers are Given in the First Column. The Percentage Falling in Each Classification is Given in the Second Column. Totals Which Do Not Equal N or 100 Per Cent Are the Result of Ommissions.

Sex	Raw Number	Per Cent of Sample
Female	125	60
Male	83	40
Marital Status		
Married	65	31
Single	145	69
Veteran Status		
Non-Veteran	119	57
Veteran	91	43
Parents' Place of Birth		
Outside United States	71	34
Inside United States	137	65
Subject's Place of Birth		
Outside United States	14	07
Inside United States	193	92
Political Preference		
Republican	44	21
Democrat	74	35
Liberal	26	12
None	47	23
Progressive	10	05
Independent	09	04
Religion		
Roman Catholic	38	18
Protestant	107	51
Jewish	29	14
None	21	10
Agnostic and Atheist	08	04
Occupation*		
Parents' Income*		
Subject's Income*		
N = 210		

* Insufficient responses.

TABLE II.—SUMMARY OF SOCIAL BACKGROUND DIFFERENCES DISCUSSION IS GIVEN IN THE TEXT*

A Republicans

	E	P	I
H	a	a	
L			−a
H			−V
L	−v		
H			d
L	−d		

B Democrats

	E	P	I
H			
L		−a	
H	v		
L			

C Progressives

	E	P	I
H			A
L	−A		
H			
L			−v
H			
L			−d

D Second Generation

	E	P	I
H		A	A
L	−A		
H	*l*	*l*	
L			−*l*

E Protestants

	E	P	I
H	a	a	
L			−a
H			+*L*
L		−*L*	
H			V
L		−V	
H			+D
L	−D		

F Roman Catholics

	E	P	I
H			v
L			

G Non-religious

	E	P	I
H			
L		D	−D

H Veterans

	E	P	I
H			a
L	−a	−a	

J Married

	E	P	I
H		*L*	
L	−*L*		

K Females

	E	P	I
H	*L*		
L			−*L*
H		D	
L	−D		

* Capital letters denote relative relationships between fields. H and L refer to high and low groups. E, P, and I refer to everyday, professional, and international fields. A, *L*, V, D, both lower case and capitals, refer to autohostile, *laissez-faire*, verbal-heterohostile, and direct-heterohostile categories.

the five per cent level. In the summary (Table II) only those data which approach significance on the five per cent level, or which are significant on the five or one per cent levels have been included. These data reinforce themselves in various proportions which may be derived from the raw data.

Republicans tend to be autohostile in everyday and professional situations, but not in international situations. They are not very verbal-heterohostile in international situations, but they are more often that way in international situations than in everyday situations. Republicans are not direct-heterohostile in everyday situations, but they are directly heterohostile in the international field. (Subtable A.)

According to the data presented in Subtable B, Democrats are verbally heterohostile in everyday situations. They are not autohostile in professional situations for the sample tested.

Progressives tend to be autohostile in international situations, but much less autohostile in the everyday field. In the international field, the Progressives are neither verbally nor directly heterohostile. (See Subtable C.)

Subtable D represents the data regarding the second-generation Americans. They do not tend to be autohostile in everyday affairs, but they are autohostile in professional and international affairs. They are *laissez-faire* in the everyday and professional fields, but not in the international field.

Protestants show autohostility in everyday and professional situations, but they are not autohostile in international situations. They are much more *laissez-faire* in international situations than in professional situations. Protestants are more verbally heterohostile in the international field than in the professional field. They are much more directly heterohostile in the international field than in the everyday field. (See Subtable E.)

Roman Catholics tend to be very verbal-heterohostile in the international field. (See Subtable F.)

Subtable G shows that the non-religious group is low in direct-heterohostility in both the professional and international fields, although it is even less directly heterohostile in the international field than in the professional field.

Subtable H concerns the veterans in this sample. They are not autohostile in the everyday and professional fields, but more autohostile in the international field. The non-veterans are less autohostile in the international field than in the everyday and professional fields.

Married people are more *laissez-faire* in professional than in everyday situations. (Subtable J.)

Women are more *laissez-faire* in the everyday field than the international field. In addition, they are more directly heterohostile in professional than everyday situations. Men react in the opposite direction from women in these two dimensions. (Subtable K.)

D. DISCUSSION

Reliability.—To summarize the discussion on reliability of the instrument used in this study, it is suggested that in some fields (e. g., everyday and professional) individuals have developed generalized responses, whereas in the international field more specific responses are forthcoming.

Laissez-faire behavior seems to be a generalized tendency for operations in all of the fields tested.

Fields and Categories.—The three fields (everyday, professional, and international) appear to be subfields of the general area of human social behavior. Significant differences do not appear between the fields. The categories (autohostile, *laissez-faire,* verbal-heterohostile, and direct-heterohostile) are highly significantly different from each other. The subgroup interaction (field-categories) is statistically significantly different. High scorers in a given category in one field tend to be high scorers in the same category in the other fields. Low scorers in a given category in one field tend to be low scorers in that category in other fields.

Social Background Data.—These data come from the high and low twenty per cent of the scorers on each field-category. All references to groups apply only to those samples which were tested in this study.

Republicans tend to be autohostile in everyday and professional situations, but contrary to this are direct-heterohostile in international situations. They tend to be more verbally heterohostile in international than everyday situations. Thus Republicans on this sample tend to direct their hostility toward themselves in everyday and professional affairs, and toward others in international affairs.

Protestants are similar to Republicans in their ways of expressing hostility. They tend to be more *laissez-faire* in the international field than in the professional field, as well as being more heterohostile in the international field than in the other two.

The Roman Catholics tested parallel the Republican and Protestant groups by their heterohostility in international affairs. This may indicate the similarity among the programs of these three groups in the international field today.

Second-generation Americans tend to be autohostile in international situations, and not very heterohostile at all in that field. They are not *laissez-faire* in the international field, but do act this way in the other fields.

The Progressives approximate the second-generation group, and like it are at variance with the Republican-Protestant-Roman Catholic group discussed earlier. Progressives tend to be very autohostile in international affairs although not in everyday affairs, and show no heterohostility in international affairs.

Veterans are highly autohostile in the international field, but not so in the other fields. They are very similar to the second-generation Americans and the Progressives. This suggests that the war experiences may have had some effect on the veterans, perhaps putting them in closer contact with international problems, making them more critical of our international policies, or perhaps motivating them towards the prevention of a future war.

The non-religious group, which shows low direct-heterohostility in the international field, lower also than in the professional field, seems to support the tendencies of the Progressives, veterans, and second-generation Americans.

The Democrats do not seem to be a homogeneous group in this sample. This may be due to the party's internal political difficulties or to regional differences. Democrats are verbal-heterohostile in everyday affairs, but not very autohostile in professional situations. Otherwise their data are not significant. These data were collected before the 1948 convention, when the party's policies were still in formulation.

Women seem to be able to express direct-heterohostility more in professional situations than are men, but men express this hostility more often in everyday situations than do women. This may indicate the need for further research regarding our cultural mores as they pertain to the ways and fields in which men and women may be expected to express their heterohostility. Perhaps male teachers and administrators are expected to act differently in frustrating situations than their female counterparts. Such data should prove to be important in the social psychological analysis of the school in our society.

Related Attitude Studies.—The social background data obtained on this inventory closely parallel attitude studies designed to measure internationalist-nationalist, war-peace, or militarism-pacifism attitudes.

Droba,[1] Hayes,[6,7] Stagner, et al,[13] and Stagner[14,15] reinforce the data in this study relative to political differences.

Eckert and Mills,[3] Katzoff and Gilliland,[8] Sappenfield,[12] Stagner et al,[13] and Stump and Lewis[17] are concerned with religious preferences which are similar to the findings in this study. Data on military training seem to differ from the data found in this study. Military training seems to be inversely related to internationalist attitudes in studies done by Dudycha,[2] H. W. Rogers,[11] Stagner, et al,[13] and Stagner.[14] Data in this study regarding subjects of foreign-born parents are reinforced by Queener's study,[10] although Eckert and Mills[3] do not find a significant difference for second-generation subjects.

Educational Implications.—Strang[16] suggests that aggression may be channelled in non-destructive directions by means of a careful educative process. She suggests that this training begin in the home and carry through the life of the individual. Queener's study[10] leads him to conclude that rewards and prestige are more frequently given for nationalistic attitudes in our culture.

Strang's solution requires action on the genetic levels of child training, and Queener's, action directed toward changes in group identification. Queener's argument lends some weight toward the concept of building international loyalties through multi- or supra-national interests such as workcamps, the arts, the sciences, etc. In any case, we seem to be weakest on action steps after we have diagnosed our difficulties.

A fruitful hypothesis is offered by Newcomb[9] who suggests that hostility of an autistic nature develops when two or more groups are not in communication with one another. With the development of this autistic hostility it becomes increasingly difficult to establish communication, and so the cycle proceeds. Postwar international relations between the United States and the Soviet Union seem to be a case in point.

The data in this study seem to support Newcomb's thesis. The recent war experiences of the veterans, their travels, their contacts with other national citizens, and their reaction to war may account for their perception of international problems. The recent arrival of the second-generation Americans' parents in this country, their probable contacts with foreign languages, foreign newspapers, recent immigrants from their home lands, and their vicarious experiences during the past war may be offered as an explanation of their perception of international affairs. The foreign policy of the Progressive Party, the newspapers, periodicals, and speeches read and listened to by Progressives may offer some clues as to the forces which form their frame of reference. We might expect some of the persons who claim no religion on this sample to be humanistic in their world outlook, and, in so doing, to avoid heterohostile reactions which might lead to war. In contrast the foreign policy of the Republican Party, the editorial policy of newspapers which Republicans might be expected to read, the traditions of individualism and isolation (physical and psychological), and the probability of less contacts between Republicans and persons of foreign birth, might account for the Republicans' frame of reference. Protestants might be expected to be similar to the Republicans in these respects. The Roman Catholic group on this sample is probably influenced greatly by the diplomatic policies of the Vatican, and so information received from that source might be expected to build their particular world outlook.

In summary, it is suggested that for this sample, the differences among expressions of hostility in the international field may be accounted for by differences in the kind and degree of verbal and physical contact (communication) with international affairs. It is further suggested that controlled changes in communication may be expected to affect the types of hostile responses of the groups studied. This is a suggestion for future research and has not been fully covered in this study.

If the process of education for better international understanding be seen as a re-educative or therapeutic one, Newcomb's hypothesis seems to be most fruitful for further action-research.

The problem becomes one of bringing the groups into better communication with the international field. The differences may then be considered to be essentially dynamic rather than genetic in character. The genetic approach to this problem is too often fraught with hypothetical causal relationships which operationally are meaningless questions. Studies in methods for the improvement of communication between groups become of greatest importance.

The problem of what type of education is needed to develop adults who express hostility in ways which are non-pathological to themselves and to society is one which deserves deepest consideration. Since individuals tend to exhibit consistent patterns of hostility, it may be important to discover the relationships between these patterns and other personality characteristics. Perhaps of greatest interest to the educator are the data demonstrating the consistency of these patterns of hostility for the individuals on this sample. This suggests that the change of an individual's expression of hostility in one field will affect a similar change of that expression in another field. Should this be demonstrated (research is underway), we would be able to educate less directly but more generally and effectively for healthy international relations involving fewer tensions.

To attempt any large scale educational program before completing further research to more conclusively determine the most effective techniques, would be presumptuous. However, the data do offer a field for hypotheses, some of which have already been raised in the discussion, and some of which appear below.

SUGGESTIONS FOR FURTHER RESEARCH

1) A study of hostility patterns might be done with subjects who are at various stages of teacher-training or post-graduate teacher-training. Such data might yield information helpful in the selection of teacher candidates. Social studies teachers might constitute a special group for study since they will be faced with the interpretation of international problems throughout their professional work.

2) The original study reported here might profitably be altered for standardization with other minority groups.

3) Subjects might be sorted into groups on the basis of their scores on this inventory, and studies made of their overt hostile

behavior in a group setting. Their reactions to the group process in education may be closely identified with their hostility patterns. Such data should prove valuable for educators using group dynamics methods.

E. CONCLUSIONS

1) Republicans, Protestants, and Roman Catholics on this sample tend to express more heterohostility in the international field than they do in the everyday and professional fields. In the everyday and professional fields the Republicans and Protestants on this sample tend to be autohostile.

2) Progressives, second-generation Americans, and veterans on this sample tend to express autohostility in the international field. Progressives and second-generation Americans tend to be much less autohostile in the everyday field. Progressives tend to be less autohostile in the professional field, too. Non-religious persons do not tend to be direct-heterohostile in the professional field, and are even less directly heterohostile in the international field.

3) With regard to the third hypothesis of this study, it has been shown for the two hundred and ten subjects tested that certain significant differences exist between extreme scorers in field-categories with reference to their social group backgrounds.

BIBLIOGRAPHY

1) D. D. Droba. "Political Parties and War Attitudes." *Journal of Abnormal and Social Psychology*, 1934, 28, 468–472.

2) George J. Dudycha. "The Attitude of College Students Toward War." *Journal of Social Psychology*, 1942, 15, 75–89.

3) R. G. Eckert and H. C. Mills. "International Attitudes and Related Academic and Social Factors." *Journal of Educational Sociology*, 1935, 9, 142–153.

4) Harry A. Grace. *A Study of the Expression of Hostility in Everyday, Professional, and International Verbal Situations.* Urbana, Ill.: Meenach, 1950.

5) ———. "Hostility, Communication, and International Tensions: I. The Hostility Inventory." (To be published)

6) Samuel P. Hayes, Jr. "The Inter-relations of Political Attitudes: I. Attitudes toward Candidates and Specific Policies." *Journal of Social Psychology*, 1937, 8, 459–482.

7) ———. "Homogeneity in Voters' Attitudes, in Relation to Their Political Affiliation, Sex, and Occupation." *Journal of Social Psychology*, 1938, 9, 141–160.

8) E. T. Katzoff and A. R. Gilliland. "Student Attitudes toward Participation in the War." *Sociometry*, 1943, 6, 149–155.

9) Theodore M. Newcomb. "Autistic Hostility and Social Reality." *Human Relations*, 1947, 1, 69–86.

10) E. L. Queener. *The Development of Internationalist Attitudes Directed toward Peace.* New Haven: Yale University, Graduate School Ph.D. Dissertation, 1947.

11) Herbert W. Rogers. "Some Attitudes of Students in the R. O. T. C." *Journal of Educational Psychology*, 1935, 26, 291–306.

12) Bert R. Sappenfield. "The Attitudes and Attitude-Estimates of Catholic, Protestant, and Jewish Students." *Journal of Social Psychology*, 1942, 16, 173–197.

13) Ross Stagner, J. F. Brown, Ralph Gundlach and R. White. "An Analysis of Social Scientists' Opinions on the Prevention of War." *Journal of Social Psychology*, 1942, 15, 381–394.

14) Ross Stagner. "Some Factors Related to Attitude toward War, 1938." *Journal of Social Psychology*, 1942, 16, 131–142.

15) ———. "Studies of Aggressive Social Attitudes: I. Measurement and Interrelation of Selected Attitudes." *Journal of Social Psychology*, 1944, 20, 109–120.

16) Ruth Strang. "Education Against Aggression." *Harvard Educational Review*, 1946, 16, 4, 273–281.

17) N. F. Stump and A. Lewis. "What Some Ministers Think about War; Some Findings from a Brief Statistical Study of Ministers' Responses toward War." *Religious Education*, 1935, 30, 135–137.

18) Frances Swineford. "Graphical and Tabular Aids for Determining Sample Size when Planning Experiments Which Involve Comparisons of Percentages." *Psychometrika*, 1946, 11, 43–49.

FACTOR ANALYSIS OF PSYCHIATRIC IMPRESSIONS

J. A. RAFFERTY and W. L. DEEMER, JR.
Department of Biometrics
USAF School of Aviation Medicine
Randolph Field, Texas

INTRODUCTION

There has always been controversy over the relative merits of personal opinions and objective measurements as selectors of personnel. In a simple situation, the grade-school teacher insists that his evaluation of a pupil has as much weight on a report card as examination scores. In a more complicated situation, the psychiatrist maintains that his concept of the etiology of a patient's disorder is more accurate than the composite results of a battery of psychological and psychiatric tests. Intuitionists and non-scientifically trained persons are often guilty of over-emphasizing the importance of personal opinions. Psychometricians and zealous immature scientists are often guilty of over-emphasizing the importance of measurements.

Few honest attempts have been made to assess the relative merits of these antipodes. Yet, it is worth while to do so. For instance, in a large scale selection program, as in the selection of young men for flying training in wartime, decisions have to be made in regard to the type of selectors used. On the one hand, impressions from interviews have the advantages of flexibility and of the weighted summarization of information due to the Gestalt-like features of the mind. On the other hand, objective test batteries have the advantages of convenient administration to large groups of subjects and statistically definable selection power. Efficient administration of such a selection program requires a knowledge of the order of magnitude of these advantages. Furthermore, it is of scientific interest to compare the statistical activities of the human brain to the techniques of psychometrics.

The work of this paper was undertaken to estimate the efficacy of clinical impressions as selectors in a particular selection program, and to study the nature of clinical impressions as products of the brain as a statistical machine. The particular program was the selection of candidates for flying training in World War II. Thus the data of this study are fixed historically, and therefore it is recognized that the results may not have as broad an application as one might hope for. The results here are presented as evidence for the conclusions; not as proof.

METHODS AND RESULTS

To a group of three hundred eighty-nine physically-fit young men inducted into the armed services in 1943, especially thorough psychiatric interviews were given by flight surgeons. The subjects were a more or less random sample of applicants for training. The interviewers were graduates of a course in aviation medicine in which the psychiatry of flying personnel was taught.[*] The psychiatric examination was called the ARMA, the Adaptability Rating for Military Aeronautics. This group of applicants was allowed to go through flying training regardless of the opinion of the psychiatrists. Hence, the actual failures in training can be compared with the predicted failures.

At each interview the psychiatrist completed a check-list of thirty-one psycho-social items and then wrote down his prediction for success in flying training (item 32) and for success in combat (item 33). An abbreviated list of these items is shown in Table I. Detailed description of the items is given elsewhere.[1] The actual outcome of elementary flying training is also given in Table I as item 34, 'Criterion.' Thus, the matrix of Table I yields the interassociations, in adjusted contingency coefficients,[†] for all pairs of items of the interview, and for each item with the criterion. Clusters of related items are evident at first glance. Valid items, those which were able to predict failures, are easily picked out in the last row or the last column where the coefficient

---

[*] A complete description of the subjects and the interviewers is given in Deemer and Rafferty.[1]

[†] Adjusted contingency coefficient is designated by C.

$$C = \sqrt{\frac{\chi^2}{N + \chi^2} \frac{\sqrt{hk}}{\sqrt{hk} - 1}}$$

where: χ^2 = Chi-square computed on h by k contingency table $N = 389$
In Table I, C is written as $100C$ to eliminate decimal points. The statistical significance of C is that of the χ^2 on which C is computed.

is prefixed by *a*, the five per cent significance level value of the coefficient (chosen arbitrarily as statistically significant association between items.) The valid items are 'Nationality,' 'Achievement,' 'Prediction for success in flying training,' and 'Prediction for success in combat.'

Clusters of items are of particular interest, since few psychosocial items *per se* were good selectors but the clinical impressions based on these items were relatively good selectors.* One would like to know how these various items are related so as to supplement one another in the considered opinion of the psychiatrists. First, a coefficient by coefficient subjective analysis was made of statistically significant associations in regard to the type of relationship between items. In Table I judgment of the type of relationship is signified by prefixing the coefficient with either *a*, for natural relationship, *b*, for systematic relationship, or *c* for semantic relationship. Furthermore, an examination of the cases of statistically insignificant association in relation to what the interviewers thought, as reflected in interassociations with the clinical impression items (nos. 31 and 32), led to the judgment of important lack of association, signified by the prefix —.

Table I. AMA Item Interassociations expressed as contingency coefficients (C's) in percent, high values of which indicate high associations.

Item		Item No.																																	
No.	Name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	Age		25	20	20	13	16	14	13	14	-15	b31	-17	b42	15	16	a22	22	24	9	12	12	b46	b44	a20	19	10	10	23	12	17	b22	-18	-16	16
2	Nationality	25		25	19	10	11	a16	5	11	a19	14	a23	-7	a25	14	8	20	10	18	b52	15	17	a14	19	16	17	0	21	19	17	a23	-17	-18	a31
3	No. Brothers	20	24		b37	16	17	7	6	a20	12	8	11	21	9	16	6	16	7	11	a27	17	11	9	5	18	10	5	a26	9	17	16	a23	a23	16
4	No. Sisters	20	19	b37		b30	b26	16	8	10	12	12	7	a23	11	14	9	14	10	15	19	15	8	10	11	15	11	9	21	13	15	20	15	17	18
5	Parental Status	13	10	18	b30		b43	a23	5	11	9	7	14	4	11	7	8	a19	11	a24	a17	6	a20	7	7	10	8	6	9	16	8	13	10	11	8
6	Upbringing	16	11	17	b26	b43		a24	4	0	6	0	6	6	2	3	7	a20	4	a15	0	5	a17	5	4	4	8	2	11	a20	4	8	9	6	9
7	Income	14	a16	7	16	a23	a24		0	0	14	11	10	15	10	0	6	6	7	4	8	12	4	2	a22	a20	a22	8	10	15	9	9	-11	-10	13
8	Hobbies	17	5	6	8	7	4	0		10	a26	a15	10	8	a17	7	2	2	1	0	11	6	5	a18	a29	a29	11	a20	a24	8	17	1	a21	a24	13
9	Questions	14	11	a29	10	11	0	8	10		7	14	4	13	a18	5	6	8	a25	a25	0	a29	2	10	8	12	a28	a22	a33	11	8	10	a27	a25	19
10	Tension	-15	a19	12	12	9	6	14	a26	7		a57	a61	a72	11	17	6	a23	0	0	13	10	7	9	12	a50	a32	a30	a34	a17	a25	14	a22	a26	13
11	Tremor	a31	14	8	12	7	0	11	a15	14	a57		a70	15	0	8	5	12	3	8	7	12	8	3	4	a40	a27	a20	17	8	15	11	a16	11	4
12	T.M.I.	-17	a23	11	7	14	6	10	10	4	a61	a79		11	10	8	15	14	12	11	0	11	5	3	10	a12	a11	a23	15	0	14	7	9	12	4
13	Achievement	b42	-7	21	a23	4	6	15	8	13	a72	15	11		17	15	13	14	15	3	10	9	b34	b27	a24	a24	16	13	a34	5	21	14	a45	a37	a30
14	Blood Sports	15	a25	9	11	11	2	10	a17	a18	11	0	10	12		a17	9	5	0	15	12	7	10	a14	a29	a29	a18	2	10	a27	a28	12	12	14	10
15	Athletics	16	14	16	14	7	3	0	7	5	17	8	8	15	a17		3	10	0	10	8	9	12	10	7	11	7	a19	a20	9	17	8	14	11	14
16	Habits	a22	8	6	9	2	7	6	2	6	6	3	13	13	6	4		5	15	0	8	8	0	2	a21	9	11	13	13	12	17	14	5	15	23
17	Alcohol	22	20	16	14	a19	a20	6	2	8	a23	12	14	14	3	18	5		8	a27	a19	12	a26	a29	0	8	12	8	15	5	4	13	a17	-14	12
18	Tobacco	24	10	7	10	11	4	7	1	a24	0	3	12	15	0	0	15	8		a27	12	a29	8	6	0	11	7	5	11	7	13	a25	9	14	14
19	Gambling	9	18	11	15	a24	a15	4	0	a25	0	8	11	2	15	10	0	a27	a27		6	17	5	3	8	10	a23	12	15	6	7	13	7	12	3
20	Religion	12	b52	a27	10	a17	0	8	11	0	15	7	0	10	12	8	0	a19	12	6		a35	a31	11	a13	10	11	5	17	8	14	19	-11	-7	13
21	Religiousness	12	15	17	15	6	5	12	6	a29	10	12	11	9	7	9	8	12	a25	17	a35		7	3	5	19	11	11	16	20	17	5	8	11	14
22	Marital	b43	17	11	8	a20	a17	4	5	2	7	8	3	b34	10	12	0	a26	8	5	a31	7		b72	a40	9	13	10	12	14	a20	11	8	10	4
23	Children	b44	a14	9	10	7	5	2	a18	10	9	3	2	b27	a14	10	2	a29	8	3	11	3	b72		a35	5	9	5	14	9	a15	13	9	5	7
24	Sex Drive	a28	19	5	11	7	4	a22	a27	8	12	4	10	a24	a20	7	a21	0	0	8	a15	5	a40	a35		a18	a18	0	18	13	a29	b24	3	0	12
25	Attitude*	19	16	18	13	10	4	a20	a29	12	a50	a40	a12	a24	a29	11	9	8	11	10	10	19	9	5	a18		a42	a33	a32	4	a28	6	a17	a26	18
26	Aggression	10	17	10	12	6	8	a22	11	a20	a12	a22	a21	16	a19	7	11	12	9	a23	11	11	13	9	a18	a42		a25	a42	a25	a27	6	a21	a22	4
27	Personality	10	0	5	9	6	2	8	a20	a22	a30	a20	a23	17	2	a19	12	8	5	12	5	11	10	5	0	a19	a33		15	21	a16	7	a21	a31	17
28	Personal Qual.	23	21	a26	22	9	11	10	a24	a35	a34	17	15	a34	18	a20	15	15	11	13	17	16	12	14	10	a32	a42	15		16	a30	13	a34	a32	19
29	Family Attitude	12	19	9	15	16	a20	15	6	11	a17	8	0	5	a27	9	12	5	7	6	8	a20	14	9	13	4	a25	21	16		b27	12	5	16	6
30	Interest in Fly	17	17	17	15	8	4	9	17	8	a25	15	14	21	a22	17	17	4	13	7	14	12	a20	b15	a29	a28	a27	a16	a30	b27		b32	a20	a23	22
31	Passenger Time	b22	a23	16	20	13	2	9	1	10	14	11	7	14	12	8	14	13	a25	13	10	5	11	13	b24	6	6	7	13	12	b32		a21	a20	13
32	Pred. Training	-18	-17	a23	15	10	5	-11	a21	a27	a22	a16	9	a45	12	14	3	a17	9	7	-11	8	8	9	3	a17	a21	a21	a34	5	a20	a21		a93	a40
33	Pred. Combat	-16	-18	a23	17	11	6	-12	a24	a25	a26	13	12	a39	14	11	13	-14	14	12	-9	11	10	5	0	a26	a22	a31	a34	16	a23	a20	a93		a41
34	Criterion*	16	a31	16	10	8	9	17	13	19	19	4	4	a30	10	14	20	12	14	3	23	14	4	7	17	18	4	17	19	6	22	17	a40	a41	

*Criterion: Eliminated for all reasons, all students in elementary training.

Note: The adjusted contingency coefficient, C, has the limiting value of 100 percent. Values of C between 20 and 25 are derived from χ^2's near the 5 percent significance level.

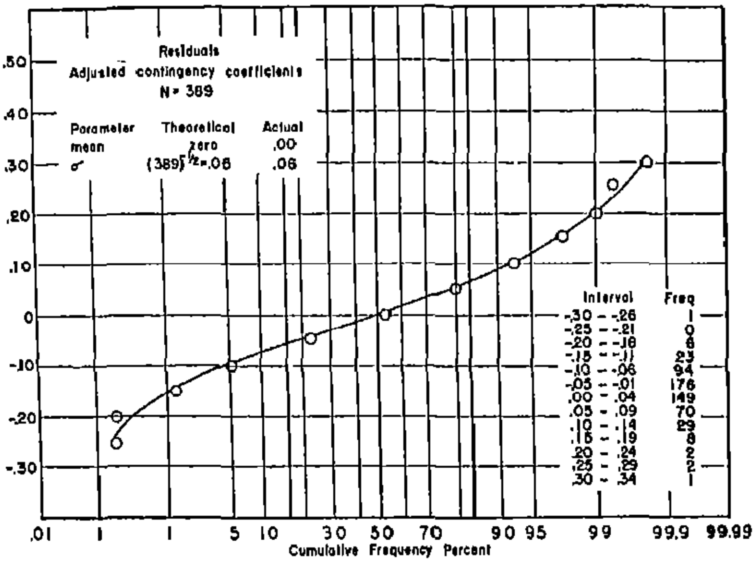
Legend:
a, b, c: Statistically significant association.
a: Natural relationship between items.
b: Systematic relationship between items.
c: Semantic relationship between items.
—: Important lack of association.

---

* The adjusted contingency coefficient, *C*, for the clinical impressions was about .40, which is comparable to the correlation coefficient of .55 for the psychological test battery given to these same subjects.[1]

Second, a multiple factor analysis was performed on the matrix of interassociations of Table I. The centroid method of factoring of Thurstone[2] was employed. Factoring was terminated when the distribution of residual contingency coefficients, after the removal of components due to the factors already found, was approximately normal with a mean of zero, and a standard deviation of nearly $(389)^{-1/2}$, the theoretical values for random fluctuations of correlation coefficients based on three hundred eighty-nine individuals. This distribution is shown in cumulative form in Figure 1.

Figure 1. Distribution of residuals in the factor analysis of ARMA items, after the extraction of four factors.



Four factors, A, B, C, and D, were sorted out as accounting for the common variance of the thirty-four items. The factor loadings and the communalities are listed in Table II, after one rotation. Other rotations, including oblique ones, were attempted with little or no improvement in the pattern, from either a practical interpretation point-of-view or a 'simple structure' point-of-view. This first rotation of 45° in AB and AC leaves the factors orthogonal, and therefore statistically independent.

TABLE II.—FACTOR LOADINGS OF ARMA ITEMS AFTER ONE ROTATION. THE FACTORS ARE ORTHOGONAL

Item		Factors				Common Variance
No.	Name	A	B	C	D	
1	Age	.502	.247	.495	.559	.8705
2	Nationality	.516	.148	.410	.240	.5137
3	No. Brothers	.368	.198	.311	.163	.2979
4	No. Sisters	.424	.141	.240	.177	.2886
5	Parental Status	.516	.007	.021	.141	.2866
6	Upbringing	.452	−.028	−.014	.092	.2137
7	Income	.255	.156	.106	.163	.1272
8	Habitus	.099	.339	.290	.332	.3190
9	Reactions	.205	.304	.198	.106	.1849
10	Tension	.092	.643	.198	.573	.7894
11	Tremor	.021	.629	.014	.636	.8008
12	V.M.I.	.071	.537	.014	.566	.6140
13	Achievement	.283	.390	.523	.389	.6618
14	Blood Sports	.283	.184	.276	.269	.2625
15	Artistic	.219	.177	.247	.276	.2165
16	Habits	.134	.191	.184	.134	.1062
17	Alcohol	.382	.113	.205	.269	.2731
18	Tobacco	.262	.134	.141	.071	.1115
19	Gambling	.325	.099	.021	.127	.1320
20	Religion	.452	.057	.339	.177	.3538
21	Religiousness	.304	.163	.113	.127	.1479
22	Marital	.573	.021	.488	.622	.9538
23	Children	.474	.049	.474	.566	.7721
24	Sex Drive	.339	.184	.431	.410	.5026
25	Attitude	.177	.516	.269	.438	.5618
26	Aggression	.191	.445	.184	.283	.3485
27	Personality	.064	.403	.134	.255	.2495
28	Personal Qual.	.233	.587	.509	.283	.7380
29	Family Attitude	.304	.163	.184	.198	.1920
30	Interest in Flying	.297	.354	.431	.346	.5190
31	Passenger Time	.325	.170	.318	.198	.2749
32	Pred. Training	.078	.686	.650	.120	.9136
33	Pred. Combat	.092	.714	.629	.099	.9237
34	Criterion	.226	.354	.452	.141	.4006
	Sum	9.538	9.481	9.485	9.546	

INTERPRETATION OF THE RESULTS

Some of the objectives of this study are not easily reached, regardless of the availability of all existing relevant data. To examine the operations of the human brain as those of a statistical machine, for instance, is not a modest endeavor. Therefore, this study is offered as a limited contribution to our knowledge of the structure and components of psychiatric impressions during interviews. Moreover, certain of the factual results are applicable in the particular setting of these data, viz., the selection of young men for flying training in the Air Force by psychiatric interview.

These factual results are more easily interpreted, so they will be discussed first. It has already been shown[1] that particular psycho-social items are of little value in selecting good flying students, with the possible exception of 'Achievement.'* Experiences of other investigators[3] with similar types of items but in quite different operational settings are in agreement with our findings. Generally speaking, one can reasonably assert that what a person has done is a fair measure of what he can do. A person's age (within physical fitness limits), religion, morality, smoking habits, number of siblings, love of parents, etc. are such weak elements in new achievements that they are easily overwhelmed by elements of motivation and specific abilities.

It is shown in Table I, however, that a psychiatrist can evaluate these various biographical and psycho-social data during an interview and arrive at a relatively high prediction coefficient. On the basis of experience and education, the interviewer subconsciously weights data and summarizes them in his impression, much the same thing that is done in the classical statistical technique of multiple regression. One might well ask how many basic factors are common to the items and the impressions. And further, one might also ask whether those basic factors account for the whole clinical impression. In conversational language, one wants to know whether "the whole is greater than its parts" (i.e., those parts listed). The evidence due to the piecemeal examination of the types of correlational relationships among items and that due to multiple factor analysis will be discussed next.

---

* 'Nationality' in Table I has as high a validity coefficient as 'Achievement.' Yet the former is not stressed in the text as important because we believe that this finding is peculiar to our sample. All the contribution to χ^2 was due to men of Slavic origin. Moreover, the χ^2 for 'Achievement' is significant at the one per cent level, whereas the χ^2 for 'Nationality' is at the four per cent level, just over the arbitrary boundary.

The psycho-social and biographical inventory represented by the first thirty-one items of Table I is quite thorough. In attaining such thoroughness one cannot help choosing items which overlap in one way or another. Some items are correlated in the physical world, although one does not know the degree of correlation ahead of time. Such statistically significant associations between items have been termed 'natural relationships between items,' and have been designated by *a* in Table I. The relationship between '24. Sex Drive' and '1. Age' is an example. Some pairs of items are correlated because the variation in one item necessarily (or at least very probably) accompanies variation in the other item. In Table I such cases of statistically significant association have been termed 'systematic relationship between items,' and have been designated by *b*. The relationship between '22. Marital Status' and '23. Number of Children' is an example. Yet other pairs of items are correlated because the meaning of one item implies part of the meaning of the other item. Again in Table I such cases have been termed 'semantic relationship between items,' and have been labelled *c*. The relationship between '10. Tension' and '12. Vaso-Motor Instability' is an example.

Finally, in the case of certain item pairs which lack statistically significant association, the opinion is made that this lack of relationship is interesting. Such cases in Table I are termed 'important lack of association,' and are labelled —. For instance, '2. Nationality' and '32. Prediction for Training' are not correlated, indicating that the psychiatrists did not consider nationality *per se* a determiner of flying ability, whereas the item '2. Nationality' in this sample of flying candidates showed significant association with the criterion; namely, pass-fail in flying training. As pointed out above, if one examines the complete contingency table for '2. Nationality' vs. '34. Criterion,' one finds that only men of Slavic origin made significant contributions to χ^2. Tracing the matter further through complete contingency tables, it is found that the psychiatrists discounted nationality *per se*, but counted the emotional stability of the subject, i.e., '10. Tension,' '11. Tremor,' and '12. Vaso-Motor

Instability'; and that men of Slavic origin in this sample (no generality implied!) of candidates for flying training showed deviations toward emotional instability as shown in their contributions to χ^2 in items 10 and 12.

Admittedly, the judgment of the type of relationship in the case of each item-pair is personal and somewhat arbitrary. However, this reduction of the number of different matters to comprehend in the perusal of Table I is helpful in arriving at simple but useful specific and general conclusions. One such specific conclusion is exhibited in the above paragraph. Other specific conclusions can be gleaned from the individual contingency tables. Reproduction of the five hundred sixty-one contingency tables here is prohibitive, although a psychiatrist interested in the factual elements of this study would find the scrutiny of these contingency tables worth while. Without the evidence of these tables, then, a few more specific conclusions will be given.

Within the age limits of Air Corps volunteers among inductees, age was not a factor in flying ability insofar as '1. Age' was not associated with '34. Criterion.' Moreover, the flight surgeons did not believe that age *per se* was a factor. Age is systematically associated with achievement (it takes time to achieve something), marital status, sexual drive, and passenger time. Yet, some of these items are valid. *Apparently, interviewers isolated age as an unimportant factor in these items.*

Even brief examination of Table I reveals a complex of items associated with the clinical predictions; viz., '8. Habitus,' '9. Reactions,' '10. Tension,' '11. Tremor,' '12. Vaso-Motor Instability,' '25. Attitude,' '26. Aggression,' '27. Personality,' and '28. Personal Qualities.' None of these items alone was strong enough to reach statistical significance. Yet, this complex allowed interviewers to extract the relevance of many items, otherwise irrelevant.

Painstaking coefficient by coefficient subjective analysis has provided some insight into the nature of psychiatric impressions in this type of setting. By educational training and by intuition (percentage admixture unknown) the interviewer seems to extract certain elementary relevancies or basic factors from the multifarious continuous and discrete sensory data offered him during the interview. He is seldom conscious (how fortunate!) of this factorization; his decision flows from some general impression of certainty in his choice among alternatives. It is not even known whether factorization occurs at some other level of the nervous system. Factorization may be a schematic fiction, poorly suited as a model for Gestalt intellectual processes. Perhaps continuous recognition in time-space-sensory data 'space' is more germane.

Assuming for the rest of this discussion that factorization is an appropriate symbolic model, one can compare brain factorization with statistical factorization. Any of a variety of systems of factor analysis can be chosen for this purpose. Here, the technique of multiple factor analysis of Thurstone[2] was applied to find the factors common to the thirty-four items. Four independent factors, A, B, C and D, were found to account for all the common variance ('communality') of the thirty-four items. The factor loadings in Table II indicate how much of each factor is in each item. In the absence of an estimate of the sampling error of factor loadings, we arbitrarily chose .30 or above as a significant loading.

Table II seen with this point-of-view suggests names of A, B, C and D. Factor A is the 'Age coefficient'—the element in those items which takes time to develop, as marital status and passenger time in commercial aircraft. Factor B is 'Poise at interview'—the aggregate of those things which convey a 'good impression' during the interview, posture and body build, reaction time, emotional stability, attitude, and others. Factor C is 'Achievement'—how the person has been able to get along in the world, as scholarship, successful married life, social acceptance, sexual adjustment. Factor D is an 'Artifact'—an apparently nonsense conglomeration, perhaps a result of sampling errors in factor loadings in this not-large sample ($N = 389$).

Factors A, B, and C correspond quite well to the factors of age, tension-tremor-aggression-complex, and achievement, uncovered in the coefficient by coefficient analysis of previous paragraphs which, in a way, was a crude factor analysis. Factor loadings in Table II show that A, 'Age coefficient,' contributed a fair share to items in general (Sum 9.538), but little to the psychiatric impressions (No. 32 .078; No. 33 .092) and little to the ability of learning to fly (No. 34 .226). Factor B, 'Poise at interview,' contributed a similar amount to all items taken together (Sum 9.481), and was important as a factor in flying training both in the opinion of interviewers (No. 32 .686; and No. 33 .714) and in actuality (No. 34 .354). Factor C, 'Achievement,' likewise contributed broadly to the variegated items (Sum 9.485), and likewise was important in opinion (No. 32 .650; No. 33 .629) and in fact (No. 34 .452).

Other interesting inferences can be drawn from the multiple factor analysis. For instance, the four isolated factors common to all items account for most of the variance in the clinical impression (No. 32 .9136; No. 33 .9237). This communality is of the same order of magnitude as the reliability of these clinical impressions, based on data from repeated interviews by the same and alternate interviewers on the same subjects. Thus, one could infer that in this case the clinical impression is 'a whole equal to the sum of its parts' (i.e. those listed). Moreover, no factor emerged which was common only to the clinical impression and the criterion. Such a doublet might have arisen had the interviewer had some sort of super-sensory insight into the factors important in flying not otherwise sampled by the existing items. Suppose one must know the command 'Open Sesame' in order to make an airplane fly, but one is not taught this command in flying training. The interviewer who determines the knowledge of this command in the subject has such insight as to produce a factor doublet in the item list under discussion. Finally, the small communality of the criterion (No. 34 .4006) implies the existence of further factors in the ability to learn to fly which are not present in all the other items nor in the clinical impressions.

SUMMARY

A multiple factor analysis was performed on the interassociation matrix of thirty-one biographical-psycho-social items, psychiatric impressions, and a pass-fail criterion in a sample of three hundred eighty-nine flying cadets. Four orthogonal factors emerged: A 'Age coefficient,' B 'Poise at interview,' C 'Achievement,' D 'Artifact.' Only Factors B and C were important in the psychiatric impressions and in the criterion, and accounted for almost all of the common variance between these two variables. All reliable variance of the psychiatric impressions was attributable to these isolated factors, indicating (not proving) that in this case the clinical impressions were 'wholes equal to the sum of their parts.'

REFERENCES

1) W. L. Deemer and J. A. Rafferty, "Experimental Evaluation of the Psychiatric Interview for Prediction of Success in Pilot Training," accepted for publication in the *J.* of *Aviation Medicine.*

2) L. L. Thurstone, *Multiple Factor Analysis,* University of Chicago Press, 1947.

3) C. W. Heath, *What People Are,* Harvard University Press, 1945.

PREDICTING SUCCESS OF ENGINEERING STUDENTS

RALPH F. BERDIE and NANCY A. SUTTER

Student Counseling Bureau
Office of the Dean of Students
University of Minnesota

Of all students graduating from high school, roughly one-fifth are seriously considering entering occupations involving scientific and technical training. Many of these students are trying to decide whether or not they should attempt to become engineers. Of a random sample consisting of five hundred men who came to the Student Counseling Bureau of the University of Minnesota for counseling, one hundred sixteen or twenty-three per cent were contemplating engineering as a career, and eighty-nine of these already were matriculated in the Institute of Technology.

In order to provide students with relevant information about their chances for success or failure in engineering training, high-school and college counselors, administrators, and teachers must have available predictive indices of known accuracy and of contemporary value.

The purpose of the study here reported was to evaluate a battery of tests assembled to assist in the prediction of academic success for students considering engineering training as one of their educational alternatives.

Several previous prediction studies for engineering students have been reported.[1,2,3,4] These studies have been of value in the counseling of students, but in light of the increasing and perhaps changing enrollment in technical schools and with concurrent curricular and professional developments, a re-appraisal of the predictive techniques should be conducted.

METHOD

In the fall quarter of 1947, a total of 1,019 freshmen were enrolled in the Institute of Technology of the University of Minnesota. During the first week of the fall quarter, all first-quarter freshmen were supposed to take a battery of tests which included:

Coöperative General Achievement Test, Part III, A Survey Test in Mathematics, Form P

Coöperative Chemistry Test, Form Q

USAFI Tests of General Educational Development, College Level, Test III, Interpretation of Reading Materials in the Natural Sciences, Form B

Revised Minnesota Paper Form Board, Series M.A.

Scores on the American Council on Education Psychological Examination, College Form, 1937 edition, and on the Coöperative English Test, Form OM, were also available for most of these students, as was the student's percentile rank in his high-school graduating class. College grades for the first quarter and for the first year were available and honor-point ratios were based on these grades.

From the five hundred twenty-four first-quarter freshmen who were tested, complete data at the end of the first quarter were avaliable for three hundred one students. The remaining two hundred twenty-three students were excluded because they had left college before the end of the quarter, they had not completed their tests, their records showed transfer credits from other colleges, or after the first week of the quarter, they had decided not to choose a major in the Institute of Technology.

Students in the Institute of Technology select their major from curricula which provide sixteen major sequences. The number of freshmen in these major groups varies. The courses these students take during their freshman year, however, are the same with one exception. Chemistry majors and chemical engineering majors take inorganic chemistry 9-10-12. All other freshmen, if they have had chemistry in high school, take inorganic chemistry 4-5-11, and if they have not had chemistry in high school, take inorganic chemistry 1-2-11.

Before students in the different major groups could be combined, the means and variances for each of the groups were compared on the basis of high-school percentile rank, American Council Examination score, and first-quarter honor-point ratio. On the basis of the consistency of the observed differences, for further analysis the group of three hundred one was divided into:

Group I—majors in chemistry and chemical engineering, all of whom had chemistry in high school. $N = 65$.

Group II—all other majors who had chemistry in high school. $N = 189$.

Group III—all other majors who did not have chemistry in high school. $N = 47$.

For each of these three groups, correlations were computed between test scores and first-quarter honor-point ratio. Multiple correlations and prediction equations were then computed and measures of statistical significance obtained.

This first part of the study was completed after the end of the fall quarter. At the end of the freshman year, a second analysis was made, based on three hundred seventy-two students. Included in this group were seventy-six cases not included in the first analysis and omitted were five who had previously been included. Not all of the tests included in the first analysis were included in the second, those tests which had no or little predictive power for fall-quarter honor-point ratio being omitted from the latter analysis. The additional cases were those students who had taken all of the tests included in the latter analysis but who had been omitted during the first analysis because some of the tests used then were not available. In the case of sixty students who had left school before the end of the year, all the grades obtained up to the time of their leaving were used in determining honor-point ratios. Of these sixty students, twenty-nine left during the first quarter, twenty-five during the second quarter, and six during the third quarter. These sixty students had mean high-school percentile ranks and honor-point ratios which were significantly lower than those for the students who remained in school. The differences between means of the two groups on the G.E.D. Test, the mathematics test, and the Paper Form Board were not significantly different.

This group of three hundred seventy-two students was treated as a whole and not broken into the curricular groups as had been done in the first analysis.

For this group, correlations were computed between first-year honor-point ratio and high-school percentile rank, Coöperative Mathematics Test, G.E.D. Test III, and the Paper Form Board. Multiple correlation coefficients and regression equations were then obtained and the significance of the beta coefficients determined. Correlations were also computed between specific test scores and grade averages in mathematics, chemistry, and drawing, for all those students for whom both the grades in each of these areas and the scores on each of these tests were available.

RESULTS

The zero order correlation coefficients between the variables, as determined at the end of the fall quarter, are presented in Table 1. These were utilized in determining the multiple correlation coefficients between all the tests and honor-point ratios for each of the three groups. These multiple coefficients were:

Group I	.79	Group II	.66	Group III	.66

All three multiple correlation coefficients were significant at the one per cent level of probability. The beta coefficients for each variable were then tested for significance and the significant ones found to be:

Group I—High-school rank significant at one per cent level G.E.D. III significant at five per cent level

Group II—High-school rank significant at one per cent level G.E.D. III significant at one per cent level, Coöperative Math. significant at one per cent level, Paper Form Board significant at five per cent level

Group III—High-school rank significant at one per cent level, G.E.D. III significant at five per cent level

For each of these groups, these variables, with the exception of the Paper Form Board, were re-combined and the following regression equations obtained:

Group I

$$Y_e = -1.0488 + .0179X_1 + .0144X_2 \qquad R_{Y_e}.X_1X_2 = .74$$

Group II

$$Y_e = -1.7211 + .0131X_1 + .0137X_2 + .0249X_3 \qquad R_{Y_e}.X_1X_2X_3 = .64$$

Group III

$$Y_e = -.9939 + .0132X_1 + .0245X_2 \qquad R_{Y_e}.X_1X_2 = .56$$

Y_e = fall-quarter HPR; X_1 = HSR; X_2 = GED; and X_3 = Coöp. Math.

The multiple R's and the beta coefficients were all significant at the one per cent level of probability.

The zero order correlations between the variables as determined at the end of the first year are presented in Table 2.

TABLE 1.—INTERCORRELATIONS BETWEEN TEST SCORES AND FALL-QUARTER HONOR-POINT RATIOS FOR EACH OF THE THREE GROUPS

Group I (chemistry and chemical engineering majors) N = 65

	HPR	HSR	GED	Math	Eng	Chem	ACE	PFB
Mean	1.30	80.31	54.55	52.38	182.92	41.63	98.19	49.89
SD	.71	16.32	11.85	8.41	41.59	13.68	23.04	6.72
HPR		.63	.60	.59	.49	.50	.53	.17
HSR			.40	.41	.38	.26	.47	−.02
GED				.52	.47	.48	.60	.22
Math					.51	.49	.64	.27
Eng						.41	.69	.20
Chem							.33	.20
ACE								.26

Group II (other majors with high-school chemistry) N = 189

	HPR	HSR	GED	Math	Eng	Chem	ACE	PFB
Mean	1.19	75.76	50.33	49.64	163.43	31.37	88.05	49.40
SD	.69	17.78	11.67	7.53	38.12	11.06	23.22	7.04
HPR		.50	.47	.48	.30	.33	.30	.29
HSR			.34	.32	.36	.28	.37	.13
GED				.45	.45	.43	.60	.34
Math					.49	.33	.50	.29
Eng						.32	.65	.27
Chem							.33	.07
ACE								.28

Group III (other majors without high-school chemistry) N = 47

	HPR	HSR	GED	Math	Eng	Chem	ACE	PFB
Mean	1.03	65.19	47.46	40.78	165.72	14.49	85.56	50.54
SD	.78	22.37	10.79	6.65	45.35	7.33	25.20	6.44
HPR		.45	.42	.34	.25	.22	.21	.21
HSR			.21	.31	.35	−.13	.37	.03
GED				.31	.42	.26	.48	−.06
Math					.33	.08	.31	.24
Eng						.02	.80	−.05
Chem							.20	.10
ACE								.01

SD = standard deviation
HPR = total honor point ratio at end of fall quarter
HSR = high school percentile rank
GED = General Educational Development Test III
Math = Cooperative Mathematics Test
Eng = Cooperative English Test
Chem = Cooperative Chemistry Test
ACE = American Council Psychological Examination

TABLE 2.—INTERCORRELATIONS BETWEEN TEST SCORES AND FIRST-YEAR HONOR-POINT RATIOS $N = 372$

	HPR	HSR	GED	Math	PFB
Mean	1.15	75.11	50.59	40.80	49.67
SD	.67	18.89	11.62	7.64	6.95
HPR		.50	.46	.41	.23
HSR			.33	.27	.10
GED				.43	.23
Math					.29

Multiple correlation coefficients were computed for several combinations of variables.

$$R_{Y_e}.X_1X_2X_3X_4 = .63$$
$$R_{Y_e}.X_1X_2X_3 = .62$$
$$R_{Y_e}.X_1X_2 = .59$$
$$R_{Y_e}.X_1X_3X_4 = .59$$
$$R_{Y_e}.X_1X_3 = .58$$
$$R_{Y_e}.X_1X_4 = .53$$

where Y_e = first year honor point ratio
X_1 = high school percentile rank
X_2 = GED Test III
X_3 = Coöperative Mathematics Test
X_4 = Paper Form Board

The regression equation resulting from the best combination of variables was computed and both the multiple correlation coefficient and the beta coefficients found significant at the one per cent level of probability.

$$Y_e = -.83785 + .01289X_1 + .01410X_2 + .01839X_3$$

Correlations between scores on separate tests and first-year grades in special areas were:

Mathematics grades with Coöperative Mathematics Test,

$$r = .43,\ N = 410$$

Chemistry grades with Coöperative Chemistry Test,

$$r = .42,\ N = 411$$

Engineering Drawing grades with Paper Form Board,

$$r = .35,\ N = 404$$

CONCLUSIONS

As has been found in previous studies, the best single predictor of over-all grade average in college was here found to be rank in high-school graduating class. This was true for three separate groups in terms of first-quarter grades and for the total group in terms of first-year grades. An almost equally efficient predictor is provided by the General Educational Development Test which predicts honor-point ratio in each group almost as well as high-school rank does. In each case, these two predictors, high-school rank and G.E.D. Test, were included in the optimal prediction battery.

In two cases, that of the fall-quarter honor-point ratio predictions for chemistry and chemical engineering majors and that of the first-year honor-point ratio prediction for the entire group, the Coöperative Mathematics Test contributed to the prediction sufficiently to warrant its inclusion in the battery.

The variables included in this study do not provide predictions of greater accuracy than those which have been reported in the past. With the exception of the G.E.D. Test, the types of tests here used are similar to those used in the past. More accurate predictions might be obtained by using additional types of tests—tests of aptitude, interest, and personality. As an effort in this direction, the freshmen entering the Institute of Technology in the fall of 1948 were given the Differential Aptitude Tests, and the results obtained with these will be reported soon.

REFERENCES

1) L. Cohen, "Predicting academic success in an engineering college and suggestions for an objective evaluation of high-school marks." *Journal of Engineering Education*, 1946, 37:381–384.

2) D. D. Feder and D. L. Adler, "Predicting the scholastic achievement of engineering students." *Journal of Engineering Education*, 1939, 29:380–385.

3) G. H. Siemens, "Forecasting the academic achievement of engineering students." *Journal of Engineering Education*, 1942, 32:617–621.

4) K. W. Vaughn, "Yale scholastic aptitude tests as predictors of success in the college of engineering." *Journal of Engineering Education*, 1944, 34:572–582.

BOOK REVIEWS

HUBERT SORENSON. *Psychology in Education.* Second Edition. New York: McGraw-Hill Book Co., 1948, pp. 535.

"Much of the first edition has been completely rewritten, some material omitted entirely, and new material added. The remainder, with the exception of a few pages, has been carefully revised. A chapter on mental health of the teacher and one on basic motives have been added." The basic plan of the second edition and apparently the author's conception of the purpose and function of educational psychology are essentially the same as the first edition.

The book contains an abundance of experimental material and is clearly and interestingly written. For these reasons it should appeal both to instructors and students. It is written more or less from the standpoint of traditional educational psychology texts; but from this standpoint it is quite up to date. It should appeal especially to instructors who take this approach to the subject. *It is definitely superior to most books of its kind.* It may prove not to make so strong an appeal to instructors who like the child development frame of reference and approach the subject from that angle.

The twenty-two chapters are distributed as follows: three, to physical, mental and social growth; four, to motivational and adjustive aspects of the subject; four, to intelligence—its measurement and correlates; six, to traditional topics of learning and memory, including individual differences, motivation, transfer and theory; one, to work and fatigue; two, to teaching methods and study skills; and two to measuring and reporting educational progress.

J. B. STROUD

State University of Iowa

HELEN SHACTER. *How Personalities Grow.* Bloomington, Ill.: McKnight and McKnight, 1949, pp. 256.

The great appeal of an earlier book designed to foster more adequate adjustment of persons at the high school age led to the present treatise. The older book has been radically revised, concepts added and new material incorporated. The discussions are no longer limited to a particular age group. They apply to adults as well as to adolescents.

The introductory material is concerned with personality patterns and likenesses and differences among personalities. This is followed by explanations of various social and sexual needs such as desire for approval, for success and the like. After noting how we frequently fail to satisfy needs and the resulting frustrations, various techniques employed for evading rather than for solving problems are explained. These include rationalization, retreat from reality, aggressive behavior, illness and others. The book ends on the theme that successful living is social living. Procedures by which one may come to understand the bases of difficulties in adjustment are outlined and suggestions given for more adequately playing the game of living.

This book, since it is planned for the individual in the general population rather than for the professional worker, is necessarily elementary. It is written in a clear and interesting style, is well organized and can be of great help to those at high school and young adult ages. Furthermore, parents and school counselors will find it very helpful in dealing with older children. It seems to the reviewer that the material is somewhat better for understanding behavior patterns in relation to personality adjustment than in furnishing hints for the readjustments needed.

MILES A. TINKER

University of Minnesota

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 41 April, 1950 Number 4

TRAITS OF PERSONALITY AND THEIR INTERCORRELATIONS AS SHOWN IN BIOGRAPHIES[1]

EDWARD L. THORNDIKE
Professor Emeritus, Teachers College, Columbia University

I have studied one or more biographies of each of the men listed below, including thirty-four writers, twenty-three statesmen, lawyers and labor leaders, fifteen business men, including two who were also inventors, nine military men, five scientists and/or inventors, three medical men, one clergyman and one monarch.

Writers: Barrie, Bennett, Browning, Burns, Byron, Carlyle, Coleridge, Cooper, Dickens, Havelock Ellis, Eugene Field, Fielding, Anatole France, Gissing, Goethe, Thomas Hardy, Hawthorne, Housman, Irving, Samuel Johnson, Lamb, Landor, Melville, Poe, Scott, Lincoln Steffens, Swinburne, Tennyson, Thackeray, Thoreau, Mark Twain, Wells, Whitman, Wordsworth.

Statesmen, lawyers and labor leaders: Bismarck, Bradlaugh, Brandeis, Carlisle Cleveland, Coolidge, Debs, Disraeli, Franklin, Alexander Hamilton, Harding, Hay, House, Andrew Jackson, Jefferson, Lincoln, Lloyd George, Ramsay Macdonald, McKinley, Tom Reed, T. Roosevelt, Carl Schurz, Webster.

Business men: John Jacob Astor, Carnegie, Jay Cooke, Eastman, Frick, Gary, Hanna, E. H. Harriman, J. J. Hill, Cyrus H. McCormick, J. Pierpont Morgan, Leland Stanford, Cornelius Vanderbilt, Wanamaker, Woolworth.

---

[1] The study reported here occupied my father during the last year of his life. He was working on it up to the time of his death. Part of the manuscript was certainly written during the week before he died and it is, as the reader will see, incomplete. It is published here without addition or change except for one or two explanatory footnotes. It represents his final legacy to psychology as an empirical, experimental science. [Robert L. Thorndike]

Military men: Grant, Haig, Hindenburg, Kitchener, Lee, Napoleon, Sherman, Washington, Wellington.

Scientists, inventors, and medical men: Darwin, Edison, Lord Kelvin, C. H. Mayo, W. J. Mayo, Newton, Osler, Pasteur.

Henry Ward Beecher, William II of Germany.

In computing the correlations I have included also ratings of nine persons well known to me, making one hundred in all.

I have rated each person for each of the traits listed below on a scale running −3, −2½, −2, −1½, −1, −½, 0, +½, +1, +1½, +2, +2½, +3 in terms of standard deviations from the median of the white males of the United States. The traits were chosen to include those reported on by Cyril Burt, some of those reported on by the author, and the 'factors' regarded as important by the Guilfords. I found myself unable to rate these men for Cattell's 'factors,' perhaps because I did not fully understand them. I did not try to use his thirty-five trait clusters because of the labor involved. Probably I should have made the effort.

Traits chosen especially for comparison with Burt: general emotionality (1), curiosity (4), worry (8), fear (6), anger (7), joy (15), sadness (9), sociability (11), sexual proclivity (17). The numbers here and later identify the traits in the correlation table (Table 1). Traits chosen especially for comparison with Thorndike: intelligence (2), craving for approval (10), dominance (12), liking for conflict (13), shyness (28), display (29), and the degree of liking for each of the following: things (30), people (31), animals (32), plants (33), children (41), words (34), reading (35), art, music, and beauty (37), bodily exercise (36), talking (42), sedentary games (45), responsibility (40), system (38), neatness (39), such indulgences as eating, drinking and smoking (43).

Traits chosen especially for comparison with Guilford factors: rathymia (19), objectivity (20), coöperativeness (21), general activity (22), inferiority complex (23), N of Guilford or nervousness (24), agreeableness (25), depression (26).

Other traits: motor ability (3), sensitiveness (5), affection (18), unselfishness (16), cruelty (14), sanity (27), relative strength of the likings for female society and for male society (44).

For special reasons I have made estimates also of (46) parental wealth at the time the man was age zero to fifteen, (47) degree to which the father or male head of the household indulged the boy, and (48) degree to which the mother or female head of the household indulged him.

When the household had no male head, I have arbitrarily entered the degree of paternal indulgence as at the median of that of the United States population.

I did not expect to discover much of value from my ratings when I began the work and my expectations grew less as it progressed. Modern biographers of the realistic school do try to include events that show the man's total personality as well as his career as a specialist, and an instructive volume could be composed of cases of trait x in person y, evidenced by y's actual behavior. But they leave many gaps.

The validity of the ratings consequently varies from very high as for sex in Byron, unselfishness in Pasteur, talking in Coleridge, agreeableness in Jay Cooke, or shyness in Theodore Roosevelt, to very low as when the only basis for rating a person in the trait is that the person was probably ordinary with respect to it because the biography gives no facts or opinions concerning that trait. This is not a safe assumption. If I had not used only men for whom recent and realistic biographies containing fairly detailed personal histories were available, many more of the ratings would have been little better than guesses. Taking them at their face value we have the intercorrelations of Table 1.

Any one of these correlations can be explored further by using only those men in whom the two traits in question have fairly dependable ratings. I intended to do this, but I cannot. My ratings were often made without written notes of the evidence supporting them. At the time it seemed sure that I would remember what it was, and my memories (at age seventy-five) are failing rapidly. Nor have I the strength to review the written notes I made when reading the biographies. So I leave the intercorrelations of Table 1 for what they may be worth, and append the ratings of the ninety-one biographies for—variables which are at least mostly better than mere chance for the use of any psychologist making similar studies. I hope the weakness of my study will not discourage them.[2]

Anybody can adduce many reasons why this way of studying

---

[2] Apparently my father intended to estimate reliabilities of the ratings for the different traits by re-rating the individuals after a period of time. Some work sheets on this aspect of the study were among his papers. However, it was not completed, and so it is impossible to select particular traits to be reported. The complete table of ratings for the ninety-one individuals and forty-eight traits is appended to this report. [R. L. T.]

TABLE 1.—INTERCORRELATIONS AMONG 48 PERSONALITY VARIABLES

	1	2	3	4	5	6	7	8
1. General emotionality		−11	−02	−04	29	02	12	00
2. Intelligence	−11		20	14	36	07	06	05
3. Motor ability	−02	20		19	00	−16	06	−11
4. Curiosity (things and people)	−04	14	19		06	−04	−04	−21
5. Sensitiveness	29	36	00	06		29	12	31
6. Fear	02	07	−16	−04	29		−14	41
7. Anger	12	06	06	−04	12	−14		08
8. Worry	00	05	−11	−21	31	41	08	
9. Sadness	01	−09	−21	−31	35	19	13	41
10. Craving for approval	11	02	08	−06	21	38	13	28
11. Sociability	23	−09	11	36	−25	−17	−06	−44
12. Dominance	08	07	17	01	−15	−35	68	−12
13. Liking for conflict	09	03	04	05	−15	−35	57	−17
14. Cruelty	−08	−10	01	−05	−09	02	47	07
15. Joy	24	−15	−03	30	−10	−18	−19	−36
16. Unselfishness	−19	16	06	22	−14	−19	−44	−09
17. Sexual proclivity	34	06	−04	−02	05	−05	18	−06
18. Affection	25	−14	−02	19	04	−15	−29	−20
19. Rathymia	35	−22	01	13	16	13	09	−04
20. Objectivity	−54	11	15	27	−32	−26	25	−10
21. Coöperativeness	−34	11	07	15	−36	−27	−41	−24
22. General activity	−07	28	35	31	−14	−27	02	−26
23. Inferiority complex	31	−24	−15	−32	27	25	−02	26

24. Nervousness	36	08	−15	−06	53	46	19	47
25. Agreeableness	30	03	11	29	−02	−01	−38	−27
26. Depression	26	−07	−30	−23	48	48	17	53
27. Sanity	−43	30	15	27	−26	−22	−17	−26
28. Shyness	−09	05	02	−30	38	27	−10	52
29. Liking for display	45	−11	−08	−02	17	17	27	−03
30. Liking for things	00	12	32	52	18	−07	−05	−14
31. Liking for people	18	−08	07	33	−26	−19	−13	−43
32. Liking for animals	16	00	22	03	−07	−12	−03	−09
33. Liking for plants	18	07	15	10	15	00	01	−07
34. Liking for words	24	37	−14	03	56	39	00	15
35. Liking for reading	12	55	−13	11	52	27	−07	22
36. Liking for bodily exercise	04	23	42	04	−02	−02	09	−12
37. Liking for art, music, beauty	17	20	02	01	31	16	−15	−09
38. Liking for system	−31	25	15	11	−03	−29	−15	−10
39. Liking for neatness	−15	21	27	−05	02	−14	−12	−12
40. Liking for responsibility	−25	15	11	18	−25	−45	−07	−25
41. Liking for children	10	07	24	−01	−11	−28	−05	−07
42. Liking to talk	41	12	−19	15	20	14	00	−15
43. Liking for eating, drinking, etc.	16	−09	−08	−04	04	19	10	−08
44. Preference of female over male society	16	−02	15	−08	39	02	08	08
45. Liking for sedentary games	−10	12	04	14	−09	−07	−20	−12
46. Parental wealth	−02	04	22	−11	02	05	17	−15
47. Paternal indulgence	09	20	−08	07	18	07	−12	−18
48. Maternal indulgence	−05	20	05	13	28	09	01	01

TABLE 1.—(*Continued*)

	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1.	01	11	23	08	09	−08	24	−19	34	25	35	−54	−34	−07
2.	−09	02	−09	07	03	−10	−15	16	06	−14	−22	11	11	28
3.	−21	08	11	17	04	01	−03	06	−04	−02	01	15	07	35
4.	−31	−06	36	01	05	−05	30	22	−02	19	13	27	15	31
5.	35	21	−25	−15	−15	−09	−10	−14	05	04	16	−32	−36	−14
6.	19	38	−17	−35	−35	02	−18	−19	−05	−15	13	−26	−27	−27
7.	13	13	−06	68	57	47	−19	−44	18	−29	09	−25	−41	02
8.	41	28	−44	−12	−17	07	−36	−09	−06	−20	−04	−10	−24	−26
9.		02	−48	−16	−16	07	−27	−12	−05	−05	−12	−26	−23	−37
10.	02		05	14	01	18	−16	−26	−08	−12	23	−27	−30	−17
11.	−48	05		16	22	01	54	05	32	21	−32	−01	24	08
12.	−16	14	16		72	52	−18	−33	20	−28	08	−10	−21	14
13.	−16	01	22	72		50	−02	−19	22	−15	07	−03	−19	17
14.	07	18	01	52	50		−23	−50	14	−52	17	−27	−45	−22
15.	−27	−16	54	−18	−02	−23		20	11	36	27	06	20	−04
16.	−12	−26	05	−33	−19	−50	20		−32	48	−33	60	59	37
17.	−05	−08	32	20	22	14	11	−32		07	28	−28	−16	−11
18.	−05	−12	21	−28	−15	−52	36	48	07		02	16	32	15
19.	−12	23	32	08	07	17	27	−33	28	02		−43	−39	−28
20.	−26	−27	−01	−10	−03	−27	06	60	−28	16	−43		67	44
21.	−23	−30	24	−21	−19	−45	20	59	−16	32	−39	67		39
22.	−37	−17	08	14	17	−22	−04	37	−11	15	−28	44	39	
23.	42	13	−20	−19	−20	17	−02	−29	13	−06	28	−37	−42	−43

24.	50	30	−25	−02	−14	17	−17	−32	15	−09	26	−49	−56	−24
25.	−36	−08	56	−26	−20	−39	51	24	13	22	22	05	29	01
26.	60	29	−34	−13	−18	18	−25	−41	17	−14	18	−57	−53	−52
27.	−43	−20	15	−04	−01	−31	12	55	−25	12	−39	59	57	44
28.	44	15	−64	−38	−40	−13	−37	10	−30	−05	−24	02	−12	−11
29.	−09	36	25	32	26	28	06	−57	42	−14	39	−54	−50	−24
30.	−04	−12	−06	−02	−14	−22	08	15	−09	27	02	18	10	21
31.	−40	−08	70	07	17	−16	56	23	12	41	23	09	29	13
32.	−11	−03	15	−03	00	−26	29	24	−15	33	−03	07	22	−02
33.	02	−03	04	05	05	−09	05	03	−07	18	−01	−01	01	10
34.	09	37	05	−07	−08	03	04	−19	21	02	28	−36	−28	−21
35.	03	00	−12	−17	−01	01	−04	03	05	−11	05	−18	−19	−04
36.	−22	04	11	09	11	−10	08	10	−12	03	−23	20	10	30
37.	−03	−02	05	−19	−19	−20	09	−10	11	18	−01	−17	01	01
38.	−11	−24	−13	−06	−04	−29	−17	43	−28	10	−62	50	52	45
39.	−12	−15	−04	−11	−09	−29	−18	28	−19	03	−48	27	39	35
40.	−21	−33	14	17	24	−28	05	45	−15	27	−47	60	60	53
41.	−16	−03	17	06	05	−21	25	32	−05	28	02	19	22	22
42.	−03	18	45	03	11	−02	38	−13	22	15	24	−34	−18	−14
43.	−16	22	31	15	08	38	12	−45	34	−21	49	−39	−41	−41
44.	20	13	−08	00	−02	14	−05	−31	26	−01	21	−19	−35	−18
45.	−25	07	33	−11	04	−16	25	30	−04	09	−03	26	28	11
46.	−19	04	−06	19	00	16	−13	−21	04	−23	01	−13	−11	−07
47.	−11	−10	−02	−11	09	−09	11	09	−04	09	−02	−05	14	08
48.	10	−05	−09	−21	02	−13	−06	−05	−09	06	01	−03	01	05

TABLE 1.—(*Continued*)

	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1.	31	36	30	26	−43	−09	45	00	18	16	18	24	12	04
2.	−24	08	03	−07	30	05	−11	12	−08	00	07	37	55	23
3.	−15	−15	11	−30	15	02	−08	32	07	22	15	−14	−13	42
4.	−32	−06	29	−23	27	−30	−02	52	33	03	10	03	11	04
5.	27	53	−02	48	−26	38	17	18	−26	−07	15	56	52	−02
6.	25	46	−01	48	−22	27	17	−07	−19	−12	00	39	27	−21
7.	−02	19	−38	17	−17	−10	27	−05	−13	−03	01	00	−07	09
8.	26	47	−27	53	−26	52	−03	−14	−43	−09	−07	15	22	−12
9.	42	50	−36	60	−43	44	−09	−04	−40	−11	02	09	03	−22
10.	13	30	−08	29	−20	15	36	−12	−08	−03	−03	37	00	04
11.	−20	−25	56	−34	15	−64	25	−06	70	15	04	05	−12	11
12.	−19	−02	−26	−13	−04	−38	32	−02	07	−03	05	−07	−17	09
13.	−20	−14	−20	−18	−01	−40	26	−14	17	00	05	−08	−01	11
14.	17	17	−39	18	−31	−13	28	−22	−16	−26	−09	03	01	−10
15.	−02	−17	51	−25	12	−37	06	08	56	29	05	04	−04	08
16.	−29	−32	24	−41	55	10	−57	15	23	24	03	−19	03	10
17.	13	15	13	17	−25	−30	42	−09	12	−15	−07	21	05	−12
18.	−06	−09	22	−14	12	−05	−14	27	41	33	18	02	−11	03
19.	28	26	22	18	−39	−24	39	02	23	−03	−01	28	05	−23
20.	−37	−49	05	−57	59	02	−54	18	09	07	−01	−36	−18	20
21.	−42	−56	29	−53	57	−12	−50	10	29	22	01	−28	−19	10
22.	−43	−24	01	−52	44	−11	−24	21	13	−02	10	−21	−04	30
23.		48	−03	57	−62	38	20	−21	−24	−10	−09	20	06	−19

24.	48		−16	73	−51	29	34	−02	−22	−21	−05	48	31	−22
25.	−03	−16		−21	13	−30	08	09	58	22	−03	12	05	12
26.	57	73	−21		−63	35	27	−11	−30	−27	−16	40	25	−31
27.	−62	−51	13	−63		−05	−50	15	25	16	−01	−18	00	24
28.	38	29	−30	35	−05		−27	−02	−54	−14	−11	07	12	05
29.	20	34	08	27	−50	−27		−06	16	−10	−03	29	04	−16
30.	−21	−02	09	−11	15	−02	−06		12	07	22	−04	−05	05
31.	−24	−22	58	−30	25	−54	16	12		23	10	03	−09	12
32.	−10	−21	22	−27	16	−14	−10	07	23		12	−09	−07	30
33.	−09	−05	−03	−16	−01	−11	−03	22	10	12		05	−08	22
34.	20	48	12	40	−18	07	29	−04	03	−09	05		60	−10
35.	06	31	05	25	00	12	04	−05	−09	−07	−08	60		02
36.	−19	−22	12	−31	24	05	−16	05	12	30	22	−10	02	
37.	−02	09	13	06	−05	04	22	09	13	09	25	26	22	10
38.	−42	−32	−06	−32	47	06	−43	16	04	11	10	−24	00	30
39.	−22	−28	12	−22	31	17	−33	−01	−02	05	06	−22	−05	41
40.	−51	−39	04	−52	53	−22	−38	15	26	10	17	−22	−16	22
41.	−11	−23	32	−33	22	−12	−21	08	31	38	09	−02	−14	24
42.	06	29	29	13	−11	−35	43	−08	38	00	06	57	36	−08
43.	13	19	00	13	−27	−24	54	−19	10	−01	−10	28	06	−16
44.	24	33	−15	25	−26	11	22	00	−13	−06	16	27	02	−02
45.	−26	−28	24	−34	28	−17	−11	03	24	13	−03	04	05	15
46.	−18	−07	−03	−09	08	−11	08	03	−14	15	11	−04	−14	12
47.	−17	−05	13	−13	21	−04	−06	17	01	25	16	09	23	04
48.	−15	04	00	00	09	18	−10	25	−05	02	08	15	15	00

TABLE 1.—(*Continued*)

	37	38	39	40	41	42	43	44	45	46	47	48
1.	17	−31	−15	−25	10	41	16	16	−10	−02	09	−05
2.	20	25	21	15	07	12	−09	−02	12	04	20	20
3.	02	15	27	11	24	−19	−08	15	04	22	−08	05
4.	01	11	−05	18	−01	15	−04	−08	14	−11	07	13
5.	31	−03	02	−25	−11	20	04	39	−09	02	18	28
6.	16	−29	−14	−45	−28	14	19	02	−07	05	07	09
7.	−15	−15	−12	−07	−05	00	10	08	−20	17	−12	01
8.	−09	−10	−12	−25	−07	−15	−08	08	−12	−15	−18	01
9.	−03	−11	−12	−21	−16	−03	−16	20	−25	−19	−11	10
10.	−02	−24	−15	−33	−03	18	22	13	07	04	−10	−05
11.	05	−13	−04	14	17	45	31	−08	33	−06	−02	−09
12.	−19	−06	−11	17	06	03	15	00	−11	19	−11	−21
13.	−19	−04	−09	24	05	11	08	−02	04	00	09	02
14.	−20	−29	−29	−28	−21	−02	38	14	−16	16	−09	−13
15.	09	−17	−18	05	25	38	12	−05	25	−13	11	−06
16.	−10	43	28	45	32	−13	−45	−31	30	−21	09	−05
17.	11	−28	−19	−15	−05	22	34	26	−04	04	−04	−09
18.	18	10	03	27	28	15	−21	−01	09	−23	09	06
19.	−01	−62	−48	−47	02	24	49	21	−03	01	−02	01
20.	−17	50	27	60	19	−34	−39	−19	26	−13	−05	−03
21.	01	52	39	60	22	−18	−41	−35	28	−11	14	01
22.	01	45	35	53	22	−14	−41	−18	11	−07	08	05
23.	−02	−42	−22	−51	−11	06	13	24	−26	−18	−17	−15

24.	09	−32	−28	−39	−23	29	19	33	−28	−07	−05	04
25.	13	−06	12	04	32	29	00	−15	24	−03	13	00
26.	06	−32	−22	−52	−33	13	13	25	−34	−09	−13	00
27.	−05	47	31	53	22	−11	−27	−26	28	08	21	09
28.	04	06	17	−22	−12	−35	−24	11	−17	−11	−04	18
29.	22	−43	−33	−38	−21	43	54	22	−11	08	−06	−10
30.	09	16	−01	15	08	−08	−19	00	03	03	17	25
31.	13	04	−02	26	31	38	10	−13	24	−14	01	−05
32.	09	11	05	10	38	00	−01	−06	13	15	25	02
33.	25	10	06	17	09	06	−10	16	−03	11	16	08
34.	26	−24	−22	−22	−02	57	28	27	04	−04	09	15
35.	22	00	−05	−16	−14	36	06	02	05	−14	23	15
36.	10	30	41	22	24	−08	−16	−02	15	12	04	00
37.		11	12	00	−08	12	12	31	−12	15	21	24
38.	11		66	64	07	−30	−48	−14	10	09	11	08
39.	12	66		46	05	−28	−46	−08	−03	18	−03	02
40.	00	64	46		30	−14	−45	−19	17	−08	02	01
41.	−08	07	05	30		−06	−22	−03	21	−07	−03	−08
42.	12	−30	−28	−14	−06		32	10	13	−16	07	07
43.	12	−48	−46	−45	−22	32		32	15	16	−01	−10
44.	31	−14	−08	−19	−03	10	32		−12	13	−07	10
45.	−12	10	−03	17	21	13	15	−12		−16	18	−02
46.	15	09	18	−08	−07	−16	16	13	−16		15	−15
47.	21	11	−03	02	−03	07	−01	−07	18	15		44
48.	24	08	02	01	−08	07	−10	10	−02	−15	44	

the problems of personality is defective, but it is at least different from the ways hitherto taken and has the merits of using life histories available to all students of behavior. Facts could be adduced to show that the methods used by Burt, the Guilfords, and Cattell are defective also. There is in fact no royal road to dependable inventories of personalities especially after early childhood. I now think that the most promising practicable way is for each of fifty or more psychologists to rate ten or more adults intimately known to him on each of an agreed upon list of traits on a scale representing the total United States white adult population. Some foundation should then pay to have two sets of intercorrelations computed for each of the fifty or more sets of measurements, one using the alleged divergences from the United States means, the other using the divergences from the actual means of each psychologist's group. The former can be combined into an average intercorrelation table (after zeta transformations) whose values are unduly raised by the fifty psychologists' errors in estimating the United States averages and variabilities. The latter can be combined into a second average intercorrelation table whose values are unduly lowered by restricted range. Taken together these two tables would, however, be far better than all that we now have on the intercorrelations of fear, anger, dominance, pugnacity, sociability, display, shyness, etc., etc. in adults.

Because the correlations of Table 1 are in part due to the large proportions of (1) writers, (2) statesmen, and (3) business men, and because there are group differences especially between the writers and others, I give in Table 2 the means for the different groups. The group differences are less than popular psychology supposes to be the case and there is always much overlapping.

I call attention to certain matters in Table 1.

It gives some support to Burt's belief that something that may be called general emotionality exists and that it is a characteristic of juvenile offenders. In our data it correlates −.11 with intellect, −.19 with unselfishness, −.54 with objectivity, −.34 with coöperativeness, −.43 with sanity, −.31 with system, −.15 with neatness, and −.25 with responsibility. It has positive correlations of .30 or more with sexual proclivity, rathymia, inferiority, N, display, and liking to talk (but also with agreeableness).

TABLE 2.*—THE MEANS OF THE GROUPS SPECIFIED IN EACH VARIABLE OF TABLE 1

	34 Writers	23 Lawyers Statesmen, etc.	15 Business Men	9 Military Men	6 Men of Science
1. General emotionality	6.9	5.9	5.7	4.5	6.6
2. Intelligence	11.0	11.4	11.1	10.4	12.0
3. Motor skill	7.7	7.7	8.7	7.9	9.2
4. Curiosity (things and people)	13.9	13.7	13.9	12.6	17.7
5. Sensitiveness	10.0	8.4	6.7	6.6	9.0
6. Fear	7.0	4.9	5.0	3.7	6.3
7. Anger	6.8	6.7	6.1	6.6	6.0
8. Worry	6.1	4.9	5.0	5.0	5.3
9. Sadness	6.4	6.4	4.9	6.2	6.0
10. Craving for approval (short scale)	5.2	4.1	3.5	4.9	4.0
11. Sociability	7.4	8.1	8.4	6.2	6.8
12. Dominance	8.2	9.4	9.0	9.6	6.8
13. Liking for conflict	6.5	7.7	7.0	6.1	4.8
14. Cruelty	5.8	4.8	4.9	4.4	3.7
15. Joy	7.4	7.6	8.0	6.1	7.8
16. Unselfishness	6.0	7.8	7.9	8.0	8.7
17. Sexual proclivity	7.2	6.9	6.9	4.7	5.7
18. Affection	7.9	8.2	8.1	8.1	8.0

TABLE 2.*—(*Continued*)

	34 Writers	23 Lawyers Statesmen, etc.	15 Business Men	9 Military Men	6 Men of Science
19. Rathymia (short scale)	5.1	3.0	2.7	1.8	3.3
20. Objectivity	6.6	9.2	9.5	10.1	9.0
21. Coöperativeness	6.3	9.2	9.7	9.7	9.8
22. General activity	7.7	9.6	10.5	9.6	11.3
23. Inferiority (short scale)	4.8	3.9	3.4	3.7	2.7
24. N of Guilford	9.2	7.0	6.4	5.7	7.7
25. Agreeableness	8.6	9.1	9.3	7.4	9.5
26. Depression	7.4	4.7	3.5	4.6	5.0
27. Sanity	7.7	9.8	10.0	9.2	10.0
28. Shyness	6.2	4.9	4.5	5.6	5.3
29. Display	8.5	6.6	7.1	4.9	5.0
30. Liking for things	6.8	6.7	7.5	6.8	10.8
31. Liking for people	7.5	8.0	8.8	6.7	7.3
32. Liking for animals	6.7	6.5	6.7	7.1	6.8
33. Liking for plants	6.8	6.5	6.5	6.4	5.8
34. Liking for words	9.9	8.5	6.4	6.3	8.0
35. Liking for reading	10.3	9.9	7.7	6.6	9.5
36. Liking for bodily exercise	6.6	6.8	7.5	7.2	5.3

37. Liking for art, music, and beauty	11.9	9.6	11.7	9.3	10.3
38. Liking for system	6.5	8.1	8.2	8.9	8.7
39. Liking for neatness	6.6	7.7	7.8	7.9	6.7
40. Liking for responsibility	6.3	9.6	9.7	9.8	9.2
41. Liking for children	6.4	7.2	7.5	7.7	7.0
42. Liking to talk	8.1	8.3	6.5	5.7	6.3
43. Liking for self-indulgence	9.3	6.9	6.8	5.2	5.3
44. Preference for women's society	6.4	4.3	4.3	5.1	3.0
45. Liking for sedentary games	6.2	7.3	7.0	5.7	6.7
46. Family wealth	9.2	8.0	8.3	10.0	9.7
47. Paternal indulgence	7.7	7.6	7.7	6.8	8.3
48. Maternal indulgence	8.3	7.7	8.0	6.3	9.0

---

* The scores on which these means are based had been converted so as to yield all positive integral values. Except as noted below, the hypothetical "median of the white males of the United States" would be represented by 7.0. In five cases the scales were either cruder because of difficulty in rating or finer because of combination of sub-scales. These variables, with the value corresponding to the population zero point, are:

4. Curiosity	12.0
10. Craving for approval	3.0
19. Rathymia	5.0
23. Inferiority	5.0
37. Liking for art, music and beauty	10.0

APPENDIX: RATINGS OF 91 BIOGRAPHEES ON 48 ATTRIBUTES

Person Rated	Attribute Rated								
	1	2	3	4	5	6	7	8	9
1. Browning	7	9	11	10	13	7	9	5	5
2. Carlyle	7	9	5	12	11	7	11	9	9
3. Wordsworth	7	7	5	12	13	9	7	9	7
4. Lamb	9	7	5	14	13	7	3	7	5
5. Hardy	5	9	11	16	13	7	5	7	5
6. Scott	7	8	9	16	11	3	11	3	4
7. Thackeray	7	7	7	14	10	7	7	5	5
8. Wells	5	9	7	17	11	7	7	7	5
9. Samuel Johnson	3	9	2	16	7	7	9	7	7
10. Gissing	9	7	5	9	10	9	7	7	7
11. Bennett	3	7	10	17	7	9	7	5	5
12. Dickens	9	7	9	17	9	7	7	5	3
13. Hawthorne	7	7	7	12	11	7	7	9	9
14. Anatole France	3	6	7	13	11	10	5	5	9
15. Washington Irving	3	5	7	14	9	7	3	7	5
16. Poe	9	6	9	13	11	9	9	7	5
17. Mark Twain	7	5	7	19	9	7	9	7	7
18. Goethe	5	9	10	10	11	7	5	3	5
19. Fielding	7	7	5	16	7	5	5	3	3
20. J. F. Cooper	5	5	7	13	9	5	7	5	7
21. Burns	7	6	7	16	9	5	9	7	9
22. Whitman	7	6	5	13	7	5	3	3	4
23. Tennyson	7	7	7	9	12	7	4	7	9
24. Swinburne	12	7	7	12	13	3	11	7	7
25. Thoreau	7	7	10	14	10	8	8	7	7
26. Coleridge	13	7	7	14	9	11	3	7	7
27. Barrie	9	6	11	12	9	5	7	7	7
28. Havelock Ellis	3	6	5	14	9	10	3	8	7
29. Byron	11	7	9	12	11	5	10	5	9
30. Housman	3	9	7	12	11	9	7	9	10
31. Melville	8	6	7	16	9	7	7	7	7
32. Landor	9	9	11	12	10	9	11	7	7
33. Eugene Field	7	5	11	14	7	7	3	3	7
34. Steffens	7	6	7	16	8	5	5	3	3
35. Webster	5	7	7	12	7	7	7	3	5
36. Bismarck	5	7	10	12	9	5	12	5	7
37. Alexander Hamilton	5	9	7	14	11	3	9	5	5
38. Cleveland	3	5	7	15	7	2	7	5	5
39. Carl Schurz	5	9	9	10	9	3	7	3	5
40. Coolidge	2	6	7	13	7	5	7	7	5
41. T. Jefferson	7	9	11	20	11	5	3	5	5
42. A. Jackson	9	6	11	16	7	3	11	5	3
43. Disraeli	7	8	7	12	9	5	5	3	5
44. Franklin	3	9	11	20	11	5	5	3	3
45. T. Roosevelt	9	8	11	15	7	4	9	5	3

APPENDIX.—(*Continued.*)

	1	2	3	4	5	6	7	8	9
46. Lincoln	5	7	9	10	9	9	5	8	11
47. McKinley	7	7	7	10	7	5	5	7	7
48. Harding	5	5	7	15	5	7	5	3	5
49. Hay	7	7	4	10	11	10	7	9	9
50. Carlisle	3	7	4	12	7	7	3	7	7
51. Brandeis	7	9	9	16	9	7	5	5	5
52. Bradlaugh	7	8	7	12	7	3	7	5	7
53. Debs	10	7	7	14	9	2	3	3	9
54. Macdonald	7	8	11	15	11	3	9	7	7
55. Lloyd George	9	8	4	16	9	4	9	3	5
56. Tom Reed	4	7	4	16	7	5	7	3	5
57. House	6	6	7	10	7	3	7	3	7
58. Frick	3	8	9	12	7	4	4	5	7
59. J. J. Hill	7	7	10	17	7	3	7	5	5
60. Hanna	7	6	9	14	7	5	7	5	3
61. Stanford	5	7	7	16	7	7	7	5	7
62. Morgan	5	8	7	10	7	3	11	7	5
63. Carnegie	8	9	9	14	9	5	3	5	3
64. E. H. Harriman	3	8	9	13	7	4	7	5	7
65. Gary	4	9	11	12	7	7	5	3	5
66. J. J. Astor	3	7	9	13	5	4	7	7	5
67. C. Vanderbilt	3	6	11	15	5	2	9	5	5
68. Jay Cooke	9	7	9	14	7	5	3	3	3
69. Wanamaker	9	7	5	16	7	7	3	3	3
70. Woolworth	7	5	5	12	5	7	5	7	5
71. Cyrus McCormick	7	5	9	14	7	7	9	5	5
72. Eastman	5	2	11	16	7	5	5	5	5
73. Napoleon	7	8	7	17	9	3	9	5	7
74. R. E. Lee	3	8	9	14	9	3	3	3	7
75. Sherman	9	7	9	12	10	4	9	9	8
76. Wellington	5	6	7	12	4	3	5	5	5
77. Kitchener	4	6	7	12	5	3	7	7	7
78. Grant	4	5	7	8	7	4	3	7	7
79. Hindenburg	1	6	7	12	3	3	5	1	5
80. Washington	3	7	9	18	7	5	7	5	3
81. Haig	5	5	9	8	5	5	11	3	7
82. Newton	1	9	11	18	11	9	9	7	7
83. Pasteur	7	8	10	18	11	7	5	7	7
84. Osler	7	8	9	18	7	5	5	3	3
85. Kelvin	7	9	11	16	7	5	5	3	7
86. Darwin	7	7	7	18	11	7	3	7	7
87. Edison	7	7	7	18	7	5	9	5	5
88. Beecher	12	5	9	9	7	4	5	7	5
89. William II	10	6	7	13	7	9	9	3	4
90. W. H. Mayo	5	7	7	12	9	7	9	5	7
91. C. J. Mayo	7	6	11	16	5	7	5	5	5

APPENDIX.—(*Continued.*)

	10	11	12	13	14	15	16	17	18	19	20	21	22
1.	6	4	9	7	5	7	5	5	9	1	5	7	11
2.	7	4	11	8	7	3	3	5	5	1	3	3	9
3.	4	4	7	5	3	5	7	9	10	3	9	7	9
4.	6	7	3	3	3	11	11	3	11	5	11	9	7
5.	5	7	8	4	3	7	7	7	9	5	11	8	9
6.	3	10	10	10	4	9	9	7	7	7	7	6	11
7.	5	10	7	7	3	9	8	7	8	9	8	7	7
8.	4	9	8	7	5	7	9	10	9	5	9	7	9
9.	6	7	11	9	7	7	7	9	7	3	7	7	3
10.	3	3	5	3	5	3	5	11	9	1	3	5	10
11.	6	10	10	10	7	7	5	8	7	1	11	7	9
12.	6	8	10	5	3	11	9	7	9	5	7	7	11
13.	5	3	5	3	3	5	7	5	9	3	6	5	3
14.	4	6	4	3	7	5	3	7	5	3	7	5	5
15.	5	9	4	4	3	13	3	7	9	7	7	9	4
16.	6	8	12	9	7	3	1	9	7	7	3	3	7
17.	7	11	9	6	7	11	7	5	7	7	5	7	7
18.	6	8	10	6	7	7	5	9	7	1	9	9	12
19.	5	11	7	7	7	9	7	10	9	9	9	9	9
20.	7	9	10	11	7	7	4	5	9	3	4	7	8
21.	5	11	9	9	9	8	9	9	9	7	7	9	9
22.	5	9	7	6	3	9	7	9	10	9	7	7	9
23.	7	5	7	3	5	5	7	3	7	1	6	7	3
24.	5	7	11	10	9	9	3	9	4	7	4	3	7
25.	5	4	8	8	7	7	5	3	7	7	7	3	9
26.	5	9	7	5	4	7	3	10	7	9	1	5	5
27.	6	7	11	7	7	7	9	7	9	9	7	5	9
28.	3	4	3	3	3	7	9	5	9	5	10	9	11
29.	5	7	10	7	11	7	1	13	7	5	3	3	5
30.	7	5	9	7	9	3	7	3	5	1	7	7	9
31.	3	7	7	5	7	9	5	5	7	5	5	5	7
32.	7	7	13	9	7	9	3	11	7	9	3	3	5
33.	3	9	7	5	9	11	9	7	11	7	7	7	5
34.	5	10	9	9	5	7	4	7	7	5	9	7	9
35.	5	9	10	9	7	5	7	7	9	7	7	9	9
36.	3	7	13	11	9	7	5	9	5	3	9	5	9
37.	3	9	11	8	5	7	5	9	7	5	7	10	11
38.	4	9	10	9	5	9	9	7	9	1	9	9	10
39.	3	7	9	9	5	7	7	7	7	1	7	9	11
40.	3	5	10	9	9	4	7	5	7	0	11	11	11
41.	3	8	9	5	1	5	11	5	11	1	9	9	11
42.	5	11	13	11	7	7	5	7	9	7	9	9	9
43.	7	9	9	8	7	7	5	8	7	9	7	7	9
44.	5	11	10	3	3	9	9	7	9	7	13	12	12
45.	5	8	11	11	5	9	6	7	8	3	9	7	13

APPENDIX.—(*Continued.*)

	10	11	12	13	14	15	16	17	18	19	20	21	22
46.	6	7	7	6	5	7	5	5	8	4	8	8	8
47.	3	10	8	4	3	7	11	7	9	0	9	11	9
48.	7	11	7	3	2	9	7	8	9	5	9	10	7
49.	5	7	7	5	3	9	7	7	7	5	9	10	7
50.	3	5	7	5	3	7	10	7	7	1	11	11	7
51.	5	9	10	11	7	7	11	7	7	1	11	10	11
52.	3	7	9	10	3	7	11	7	9	1	11	9	11
53.	3	7	7	7	1	9	11	7	11	3	7	9	11
54.	6	7	9	7	3	7	5	7	9	1	11	9	11
55.	3	9	11	11	5	11	7	7	9	1	9	9	7
56.	3	8	9	9	9	11	9	5	9	1	10	7	7
57.	1	7	9	5	4	7	9	7	7	1	10	12	9
58.	3	7	7	5	3	7	7	7	7	1	9	11	9
59.	3	8	9	9	7	7	9	7	9	5	11	11	12
60.	4	11	11	9	3	9	9	7	9	5	9	11	11
61.	4	9	8	7	5	7	7	7	9	5	9	11	9
62.	2	7	12	8	7	7	5	11	7	1	9	10	7
63.	5	10	9	7	2	9	11	5	11	5	9	12	13
64.	3	6	11	9	7	7	9	7	9	1	11	7	11
65.	3	10	7	6	3	7	9	7	7	1	9	13	11
66.	5	9	9	5	7	9	7	7	9	3	9	7	9
67.	3	7	12	11	7	7	5	10	5	5	11	7	11
68.	3	10	7	5	3	12	9	7	9	2	9	13	11
69.	3	11	7	5	1	11	9	7	9	1	9	11	13
70.	3	7	8	7	7	7	5	7	7	3	9	7	9
71.	5	7	11	11	7	7	7	4	7	3	9	4	11
72.	3	7	7	5	5	7	11	4	7	0	11	11	11
73.	5	7	13	9	7	7	7	5	7	0	11	9	13
74.	3	7	4	2	1	7	11	4	9	1	9	9	10
75.	7	9	10	7	2	7	11	7	11	3	9	9	11
76.	5	7	10	7	5	7	5	7	7	3	11	11	9
77.	3	2	11	5	3	5	7	3	7	1	11	9	11
78.	4	4	9	4	5	7	9	3	7	5	9	9	3
79.	5	7	9	7	7	7	9	3	7	1	10	10	10
80.	7	8	11	7	7	3	9	7	9	1	11	10	10
81.	5	5	9	7	3	5	7	3	9	1	10	11	9
82.	6	3	9	2	7	5	5	3	3	5	9	9	11
83.	5	4	5	5	1	7	13	7	11	1	11	11	13
84.	4	10	7	7	3	12	11	7	9	5	7	11	11
85.	3	9	7	3	3	9	7	7	9	1	9	11	11
86.	3	8	3	1	1	9	11	5	9	1	11	10	9
87.	3	7	10	7	7	5	5	5	7	7	7	7	13
88.	7	11	9	5	7	7	3	11	7	5	3	7	9
89.	7	10	13	5	7	7	3	5	7	5	4	4	9
90.	7	7	9	8	5	7	9	7	9	0	11	11	12
91.	5	11	5	4	3	11	9	7	10	2	11	11	11

APPENDIX.—(*Continued.*)

	10	11	12	13	14	15	16	17	18	19	20	21	22
1.	6	4	9	7	5	7	5	5	9	1	5	7	11
2.	7	4	11	8	7	3	3	5	5	1	3	3	9
3.	4	4	7	5	3	5	7	9	10	3	9	7	9
4.	6	7	3	3	3	11	11	3	11	5	11	9	7
5.	5	7	8	4	3	7	7	7	9	5	11	8	9
6.	3	10	10	10	4	9	9	7	7	7	7	5	11
7.	5	10	7	7	3	9	8	7	8	9	8	7	7
8.	4	9	8	7	5	7	9	10	9	5	9	7	9
9.	6	7	11	9	7	7	7	9	7	3	7	7	3
10.	3	3	5	3	5	3	5	11	9	1	3	5	10
11.	6	10	10	10	7	7	5	8	7	1	11	7	9
12.	6	8	10	5	3	11	9	7	9	5	7	7	11
13.	5	3	5	3	3	5	7	5	9	3	5	5	3
14.	4	6	4	3	7	5	3	7	5	3	7	5	5
15.	5	9	4	4	3	13	3	7	9	7	7	9	4
16.	6	8	12	9	7	3	1	9	7	7	3	3	7
17.	7	11	9	6	7	11	7	5	7	7	5	7	7
18.	6	8	10	6	7	7	5	9	7	1	9	9	12
19.	5	11	7	7	7	9	7	10	9	9	9	9	9
20.	7	9	10	11	7	7	4	5	9	3	4	7	8
21.	5	11	9	9	9	8	9	9	9	7	7	9	9
22.	5	9	7	6	3	9	7	9	10	9	7	7	9
23.	7	5	7	3	5	5	7	3	7	1	6	7	3
24.	5	7	11	10	9	9	3	9	4	7	4	3	7
25.	5	4	8	8	7	7	5	3	7	7	7	3	9
26.	5	9	7	5	4	7	3	10	7	9	1	5	5
27.	6	7	11	7	7	7	9	7	9	9	7	5	9
28.	3	4	3	3	3	7	9	5	9	5	10	9	11
29.	5	7	10	7	11	7	1	13	7	5	3	3	5
30.	7	5	9	7	9	3	7	3	5	1	7	7	9
31.	3	7	7	5	7	9	5	5	7	5	5	5	7
32.	7	7	13	9	7	9	3	11	7	9	3	3	5
33.	3	9	7	5	9	11	9	7	11	7	7	7	5
34.	5	10	9	9	5	7	4	7	7	5	9	7	9
35.	5	9	10	9	7	5	7	7	9	7	7	9	9
36.	3	7	13	11	9	7	5	9	5	3	9	5	9
37.	3	9	11	8	5	7	5	9	7	5	7	10	11
38.	4	9	10	9	5	9	9	7	9	1	9	9	10
39.	3	7	9	9	5	7	7	7	7	1	7	9	11
40.	3	5	10	9	9	4	7	5	7	0	11	11	11
41.	3	8	9	5	1	5	11	5	11	1	9	9	11
42.	5	11	13	11	7	7	5	7	9	7	9	9	9
43.	7	9	9	8	7	7	5	8	7	9	7	7	9
44.	5	11	10	3	3	9	9	7	9	7	13	12	12
45.	5	8	11	11	5	9	6	7	8	3	9	7	13

APPENDIX.—(*Continued.*)

	10	11	12	13	14	15	16	17	18	19	20	21	22
46.	6	7	7	6	5	7	5	5	8	4	8	8	8
47.	3	10	8	4	3	7	11	7	9	0	9	11	9
48.	7	11	7	3	2	9	7	8	9	5	9	10	7
49.	5	7	7	5	3	9	7	7	7	5	9	10	7
50.	3	5	7	5	3	7	10	7	7	1	11	11	7
51.	5	9	10	11	7	7	11	7	7	1	11	10	11
52.	3	7	9	10	3	7	11	7	9	1	11	9	11
53.	3	7	7	7	1	9	11	7	11	3	7	9	11
54.	6	7	9	7	3	7	5	7	9	1	11	9	11
55.	3	9	11	11	5	11	7	7	9	1	9	9	7
56.	3	8	9	9	9	11	9	5	9	1	10	7	7
57.	1	7	9	5	4	7	9	7	7	1	10	12	9
58.	3	7	7	5	3	7	7	7	7	1	9	11	9
59.	3	8	9	9	7	7	9	7	9	5	11	11	12
60.	4	11	11	9	3	9	9	7	9	5	9	11	11
61.	4	9	8	7	5	7	7	7	9	5	9	11	9
62.	2	7	12	8	7	7	5	11	7	1	9	10	7
63.	5	10	9	7	2	9	11	5	11	5	9	12	13
64.	3	6	11	9	7	7	9	7	9	1	11	7	11
65.	3	10	7	6	3	7	9	7	7	1	9	13	11
66.	5	9	9	5	7	9	7	7	9	3	9	7	9
67.	3	7	12	11	7	7	5	10	5	5	11	7	11
68.	3	10	7	5	3	12	9	7	9	2	9	13	11
69.	3	11	7	5	1	11	9	7	9	1	9	11	13
70.	3	7	8	7	7	7	5	7	7	3	9	7	9
71.	5	7	11	11	7	7	7	4	7	3	9	4	11
72.	3	7	7	5	5	7	11	4	7	0	11	11	11
73.	5	7	13	9	7	7	7	5	7	0	11	9	13
74.	3	7	4	2	1	7	11	4	9	1	9	9	10
75.	7	9	10	7	2	7	11	7	11	3	9	9	11
76.	5	7	10	7	5	7	5	7	7	3	11	11	9
77.	3	2	11	5	3	5	7	3	7	1	11	9	11
78.	4	4	9	4	5	7	9	3	7	5	9	9	3
79.	5	7	9	7	7	7	9	3	7	1	10	10	10
80.	7	8	11	7	7	3	9	7	9	1	11	10	10
81.	5	5	9	7	3	5	7	3	9	1	10	11	9
82.	6	3	9	2	7	5	5	3	3	5	9	9	11
83.	5	4	5	5	1	7	13	7	11	1	11	11	13
84.	4	10	7	7	3	12	11	7	9	5	7	11	11
85.	3	9	7	3	3	9	7	7	9	1	9	11	11
86.	3	8	3	1	1	9	11	5	9	1	11	10	9
87.	3	7	10	7	7	5	5	5	7	7	7	7	13
88.	7	11	9	5	7	7	3	11	7	5	3	7	9
89.	7	10	13	5	7	7	3	5	7	5	4	4	9
90.	7	7	9	8	5	7	9	7	9	0	11	11	12
91.	5	11	5	4	3	11	9	7	10	2	11	11	11

APPENDIX.—(*Continued.*)

	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1.	3	9	9	5	11	11	7	7	7	11	7	13	13	11
2.	3	11	3	11	11	7	5	3	3	7	5	11	13	7
3.	3	11	7	7	9	7	7	11	7	7	11	11	11	9
4.	7	9	11	7	5	9	7	7	9	7	7	13	13	9
5.	5	7	9	5	11	7	5	9	7	7	7	11	11	7
6.	3	5	11	3	11	3	9	7	10	9	7	9	11	11
7.	3	7	10	5	11	7	7	7	9	7	3	9	11	5
8.	5	7	9	5	11	7	7	9	7	5	7	11	13	7
9.	3	13	7	13	11	3	9	9	11	5	3	13	13	2
10.	7	11	7	13	5	9	7	5	3	3	3	9	11	3
11.	1	9	9	7	11	5	9	7	11	3	5	9	9	7
12.	3	10	10	7	10	5	11	7	11	9	5	9	9	9
13.	3	9	9	9	7	10	7	7	7	7	7	9	9	9
14.	7	11	7	9	7	9	9	7	3	3	7	11	11	3
15.	3	7	11	7	9	3	9	7	9	7	7	9	9	7
16.	3	7	7	11	1	3	13	7	7	5	7	9	9	5
17.	3	9	10	5	11	5	11	7	11	7	5	11	9	3
18.	3	7	7	5	11	3	9	9	7	5	9	13	11	9
19.	5	5	11	5	11	5	7	5	11	5	3	9	7	5
20.	1	5	7	3	7	3	11	8	7	9	7	7	7	9
21.	5	9	7	10	4	5	7	5	9	7	7	8	10	7
22.	5	8	11	5	5	3	11	3	9	5	5	9	10	3
23.	7	11	9	11	7	12	9	7	7	7	7	11	9	5
24.	9	13	5	9	3	5	11	5	5	5	5	11	13	8
25.	9	10	4	9	7	7	7	9	5	7	9	7	11	7
26.	7	13	9	11	3	5	11	7	9	7	5	13	13	5
27.	7	11	7	9	9	9	5	7	7	7	5	11	9	7
28.	7	10	9	7	9	10	5	3	5	7	5	9	11	5
29.	11	12	7	9	3	10	13	5	5	7	5	7	9	9
30.	5	12	5	9	5	9	7	5	3	5	9	9	9	7
31.	5	11	7	9	5	5	9	9	7	7	7	7	11	7
32.	7	9	9	9	4	3	11	7	7	9	11	13	11	7
33.	5	9	9	5	7	3	5	7	10	10	7	9	7	2
34.	1	5	13	3	9	3	11	7	9	9	3	7	7	7
35.	3	7	9	5	9	3	9	7	9	5	7	11	11	5
36.	3	9	7	7	9	3	9	7	7	7	7	7	9	9
37.	1	9	9	5	9	1	5	5	9	5	9	9	11	7
38.	3	5	7	3	10	4	3	5	7	5	5	4	7	3
39.	3	5	7	3	10	5	9	5	7	5	5	11	13	3
40.	7	7	7	7	11	11	3	5	7	5	3	9	11	7
41.	3	7	9	3	12	7	5	13	9	9	11	9	13	9
42.	3	7	9	3	9	2	7	7	9	9	9	9	3	7
43.	3	9	7	3	9	3	9	5	7	5	11	11	9	5
44.	2	7	13	1	13	3	7	13	11	7	7	11	11	9
45.	2	7	8	4	10	3	10	8	7	9	5	8	12	12

APPENDIX.—(*Continued.*)

	23	24	25	26	27	28	29	30	31	32	33	34	35	36
46.	9	10	10	11	4	7	5	6	8	6	3	11	9	6
47.	3	7	11	5	11	7	7	5	9	7	7	7	7	7
48.	7	7	12	5	9	3	9	7	9	9	5	9	5	7
49.	7	11	9	11	9	11	3	5	5	5	5	9	9	7
50.	3	2	9	3	11	7	5	5	5	5	5	9	11	2
51.	3	7	11	3	11	3	7	7	7	7	5	7	11	5
52.	3	7	9	5	11	3	7	7	9	7	7	9	11	7
53.	3	9	12	5	9	5	9	7	11	7	7	7	11	7
54.	5	9	7	7	9	11	9	7	7	5	5	9	11	9
55.	7	4	10	3	9	3	10	7	9	11	10	7	11	9
56.	3	5	7	3	11	3	3	7	7	5	7	10	12	8
57.	3	3	10	3	11	4	1	5	9	5	5	3	9	7
58.	3	7	9	3	12	7	3	7	7	7	5	7	9	9
59.	3	5	10	3	11	3	3	7	9	7	7	7	12	9
60.	3	5	11	3	11	3	10	7	11	7	7	7	7	5
61.	3	7	9	5	9	3	7	7	9	9	7	5	7	7
62.	3	7	7	4	9	3	11	5	5	9	3	7	7	2
63.	1	7	9	3	13	4	7	9	11	7	7	7	11	7
64.	3	9	5	3	11	7	3	9	7	7	5	5	7	9
65.	3	4	11	3	11	5	4	7	9	7	7	7	10	10
66.	5	7	10	5	9	5	7	7	10	5	5	7	7	9
67.	3	5	7	3	7	3	9	9	7	3	5	4	3	9
68.	3	5	13	3	11	3	8	7	11	7	7	7	7	9
69.	3	7	13	3	11	3	7	8	11	9	7	11	9	9
70.	7	9	9	5	5	3	12	7	11	7	5	5	7	3
71.	5	7	7	3	9	7	9	7	7	5	10	5	5	7
72.	3	5	9	3	11	9	3	9	7	5	11	5	7	9
73.	1	9	7	5	7	3	7	7	7	5	7	11	11	7
74.	7	7	11	5	9	5	3	7	7	7	7	7	7	7
75.	5	11	7	7	7	8	7	7	7	9	5	9	7	7
76.	3	3	7	3	7	4	5	5	7	7	9	7	7	7
77.	3	7	7	3	11	9	2	7	4	7	10	3	3	7
78.	7	5	7	7	7	6	5	7	5	11	3	5	9	3
79.	1	1	7	3	13	5	4	7	9	7	5	5	5	9
80.	3	5	9	5	11	3	8	7	7	7	5	7	7	7
81.	3	3	5	3	11	7	3	7	7	9	7	3	3	11
82.	3	9	7	7	9	9	7	13	3	5	3	7	9	3
83.	3	9	11	7	9	9	5	13	7	7	7	9	9	5
84.	3	5	11	3	12	3	3	9	11	7	7	9	9	5
85.	3	7	10	3	9	3	7	9	9	9	5	9	9	11
86.	3	7	11	7	12	5	1	8	7	8	7	7	10	5
87.	1	9	7	3	9	3	7	13	7	5	5	7	11	3
88.	7	7	9	7	7	5	12	7	9	5	9	9	7	7
89.	5	11	9	7	7	3	13	5	7	3	5	7	7	7
90.	3	5	7	3	11	7	7	7	7	7	5	7	7	7
91.	3	3	11	3	11	4	3	9	9	12	9	5	7	9

APPENDIX.—(*Continued.*)

	37	38	39	40	41	42	43	44	45	46	47	48
1.	18	9	9	7	5	9	9	11	5	13	13	13
2.	11	9	7	7	5	9	5	5	5	7	9	11
3.	12	7	7	9	7	7	5	7	5	11	9	9
4.	12	9	7	7	7	11	9	5	11	5	7	7
5.	16	7	7	8	7	5	7	8	5	10	7	7
6.	11	9	9	9	9	9	11	2	7	11	7	9
7.	11	3	7	7	7	9	11	8	7	11	7	9
8.	12	7	5	7	7	9	9	5	7	7	8	11
9.	9	7	3	7	3	13	11	3	7	8	7	7
10.	9	5	11	5	5	7	7	7	5	8	7	7
11.	13	11	11	13	5	7	11	7	7	10	5	7
12.	12	9	7	9	11	7	7	5	7	8	9	5
13.	13	7	7	7	9	7	7	5	5	11	7	11
14.	13	5	3	3	2	9	9	7	5	9	9	13
15.	14	5	7	7	5	9	9	7	5	9	9	9
16.	13	5	7	3	5	7	11	9	5	9	5	10
17.	8	3	3	3	7	11	11	7	7	7	7	7
18.	14	9	7	9	9	11	12	10	7	11	10	9
19.	10	5	5	5	9	7	11	5	7	9	7	7
20.	13	9	7	9	9	8	7	5	11	11	9	7
21.	14	7	7	7	7	5	9	7	7	5	7	7
22.	11	4	3	3	4	9	10	3	5	8	7	7
23.	14	9	9	5	8	7	9	7	7	10	9	9
24.	10	7	9	7	9	9	5	7	5	11	7	7
25.	11	7	5	3	7	7	9	7	5	8	9	9
26.	11	3	5	1	5	13	10	3	5	8	11	9
27.	7	4	5	7	11	5	9	9	7	7	7	9
28.	14	9	7	7	3	7	9	8	5	9	7	7
29.	10	4	7	3	3	7	13	9	5	11	7	3
30.	9	7	9	9	3	2	9	3	5	8	7	9
31.	11	7	7	4	3	9	10	7	7	9	7	7
32.	13	3	5	3	9	9	11	7	5	13	5	5
33.	12	5	5	7	7	9	11	7	5	11	7	9
34.	13	5	7	7	7	7	13	5	7	10	9	9
35.	10	7	7	9	9	9	11	3	9	9	11	7
36.	7	7	5	9	7	7	12	7	7	13	9	5
37.	10	11	9	11	7	7	7	5	7	7	7	7
38.	4	9	7	9	7	7	9	1	7	7	7	7
39.	14	9	9	11	7	9	5	1	7	9	9	9
40.	7	9	9	11	7	3	3	1	3	9	7	7
41.	13	12	12	11	7	7	3	3	7	11	7	7
42.	4	7	7	11	9	9	11	7	9	9	7	9
43.	11	7	7	9	5	11	11	10	5	12	9	9
44.	12	9	9	11	9	9	9	8	11	9	7	7
45.	10	7	7	11	7	10	9	3	6	11	7	7

APPENDIX.—(*Continued.*)

	37	38	39	40	41	42	43	44	45	46	47	48
46.	6	4	6	8	10	9	4	3	7	3	5	7
47.	10	9	9	10	7	9	7	3	7	9	7	7
48.	9	5	7	7	5	11	11	3	11	7	7	9
49.	12	7	9	7	7	5	5	5	7	9	9	9
50.	10	7	7	11	7	7	9	2	9	3	5	5
51.	10	9	7	9	7	9	5	7	9	10	9	9
52.	10	9	9	11	7	9	3	7	5	7	7	5
53.	10	7	7	9	9	11	5	6	7	4	7	9
54.	11	9	9	10	7	9	3	9	7	2	3	11
55.	13	5	5	9	7	9	7	3	5	4	12	10
56.	7	5	5	9	7	11	7	2	11	8	9	7
57.	10	12	10	9	5	5	3	2	5	12	9	7
58.	15	11	9	11	7	5	9	7	11	9	9	9
59.	13	7	7	9	9	7	5	3	7	7	9	7
60.	9	7	7	11	7	7	5	2	11	11	9	7
61.	10	7	7	9	7	7	5	2	7	9	9	9
62.	14	7	7	8	9	4	12	9	5	13	7	7
63.	13	9	9	12	7	9	7	3	7	5	11	11
64.	7	7	5	9	9	5	7	6	7	6	7	7
65.	14	9	11	11	7	7	5	3	7	9	8	10
66.	12	7	7	9	10	7	5	3	3	7	3	7
67.	7	7	7	9	4	4	7	7	9	7	7	9
68.	12	9	7	9	9	7	9	3	9	11	9	7
69.	12	11	9	12	9	10	5	2	5	7	7	7
70.	14	9	7	9	7	7	9	5	5	7	7	9
71.	9	7	9	9	7	7	5	6	5	9	7	7
72.	15	9	9	9	5	6	7	5	7	7	7	7
73.	11	9	7	11	5	11	5	5	7	10	7	7
74.	10	9	11	9	11	7	3	11	5	9	7	7
75.	11	7	5	9	9	13	7	5	7	7	7	7
76.	11	9	7	12	9	5	7	8	5	11	7	7
77.	11	7	7	11	9	2	3	3	5	11	7	7
78.	6	7	5	7	7	3	9	3	7	9	7	3
79.	10	9	9	9	7	5	5	5	7	11	7	7
80.	10	12	9	11	5	3	5	4	5	11	7	5
81.	7	11	11	9	7	2	3	2	3	11	5	7
82.	7	7	7	3	5	5	5	1	3	11	7	9
83.	12	11	9	11	7	3	1	3	7	9	11	9
84.	10	7	9	11	9	9	3	1	7	9	9	9
85.	12	9	9	9	7	9	9	3	7	10	7	7
86.	10	11	9	11	7	7	5	7	9	11	7	7
87.	11	7	3	10	7	5	9	3	7	8	9	11
88.	12	5	5	5	7	9	11	7	7	9	5	3
89.	13	3	7	5	4	10	11	2	3	13	5	2
90.	11	11	11	9	7	7	5	2	9	9	9	9
91.	11	7	7	7	10	5	5	3	9	9	9	9

It gives support to the hypothesis that fear and anger are independent gene-caused traits in man. The correlation is −.14; and Burt found a similar slight negative correlation as an exception to his general view of emotional linkages.[3]

[FOOTNOTE FOR APPENDIX]

In the ratings given here, with certain exceptions which will be noted, the integers correspond to half standard deviation units of the distribution of white males in the United States, while the median of this distribution is represented by the number 7. The exceptions are as follows:

Variable 4. Curiosity. This is a composite of two separate ratings. It is expressed so that the neutral point is 12, and its standard deviation in the population would be more than 2.

Variable 37. Liking for art, music, beauty. Also a composite. Indifference point 10.0, and standard deviation greater than 2.

Variable 10. Craving for approval. The number scale has been shifted so that the presumed population median for this variable is 3.

Variable 19. Rathymia. Number scale shifted so that presumed population median is 5.

Variable 23. Inferiority. Number scale shifted so that presumed population median is 5.

---

[3] At this point the manuscript comes to an end. It is clearly unfinished, but I do not know on what further aspects of the data my father intended to comment. The results are now available to all for such further analysis and interpretation as they may care to make. [R. L. T.]

CARL E. SEASHORE

1866–1949

Experimental methods were introduced in the study of human behavior during the latter part of the Nineteenth Century. The first psychological laboratory was established in 1879 by Wilhelm Wundt at the University of Leipzig. Immediately thereafter American universities followed with the founding of a score of psychological laboratories before the end of the century.

During the next fifty years psychology as a science grew from an infant to a husky youth. At first emphasis was largely on sensory and perceptual processes as measured by motor responses and reaction times. Very soon, however, the field broadened. Interests of investigators expanded enormously to cover a wide variety of aspects of human behavior and social relations: learning and growth, aptitudes and interests, motivations and incentives, attitudes and opinions, intelligence and personality. Ventures into these new fields called for refinements in method and particularly the application of statistical techniques.

Furthermore this first half of the Twentieth Century is also the period during which American universities overtook European centers of learning and rose to world leadership in research.

Carl E. Seashore's adult life span coincided with this half century of the flowering of research in American universities. The uniqueness of his contribution is the combination of outstanding achievement and leadership (1) not only in his own field of psychology as an investigator and guide of the graduate work of numerous students, but (2) also in the over-all field of graduate research at the University of Iowa in setting high standards of attainment.

For forty years he was director of the psychological laboratory and for thirty of these years he was also dean of the graduate college followed by a decade of semi-retirement devoted to research and writing. Coming almost immediately, after his graduate studies at Yale, to Iowa in 1897, his half century of teaching, research and administrative leadership is unequalled at the University of Iowa and probably seldom matched anywhere else.

Dr. Seashore came from sturdy pioneer stock and carried these same qualities into his pioneering in psychology. He was born in Morlunda, Sweden, January 28, 1866. When three years old his family moved to the United States and settled in Iowa. He was graduated from Gustavus Adolphus College, St. Peter, Minnesota, in 1891. He received his doctorate at Yale in 1895 and after two years as assistant at Yale, he came to the University of Iowa to take charge of the psychological laboratory whose beginnings had been made by Professor George T. W. Patrick.

In his early years he invented the Voice Tonoscope, the Audiometer, and the Sound Perimeter. Among his chief contributions to psychological science was his development of a system of analysis and measurement of musical aptitude as a basis for vocational guidance in music. Another of his major achievements was the establishment of the Child Welfare Research Station.

Through Dean Seashore's vision and perseverance, the Iowa psychological laboratory became one of a small handful of top-ranking laboratories, both in research contributions and in training outstanding professional leaders.

He was a member of Sigma Xi and a fellow of the American Psychological Association of which he was president in 1911. In 1920 he was chairman of the division of Psychology and Anthropology of the National Research Council. He received many honors and recognitions. In 1948 the Board of Education of the state of Iowa conferred upon him the title of "Distinguished Service Professor."

Dr. Seashore died on October 16, 1949, only two months after his wife had passed away. Always closely tied to each other and supplementing each other's interests, their home and family were their delight and satisfaction. Only those who were privileged to come under his guidance and sense the philosophy that permeated his work and his home, can fully appreciate his worth and influence.

DANIEL STARCH

Daniel Starch and Staff
New York 17, New York

NOTES ON THE MEASUREMENT OF MENTAL SPEED

MERLE W. TATE
University of Pennsylvania

Mental speed or rate of performing intellectual tasks has long been of interest to psychologists. The two most persistent issues—and they are related—have been whether mental speed is independent of difficulty and type of task and whether it is indicative of profundity or altitude of intellect, accuracy or quality of work, temperamental factors, achievement in school, and vocational success.

The inconsistencies and contradictions in the voluminous literature suggest either that speed is an unstable dimension of intellect or that inappropriate units and methods of measuring it have frequently been employed. Regarding the latter possibility, it is easy to show that the findings in a number of studies may have been consequent to the method of investigation, rather than to the rôle of speed. Substantial correlation of scores obtained from timed and untimed administration of group tests, in view of the part-whole circumstance, is inevitable—whatever the relation of speed to altitude. The meaning of the significant inverse relationship between cumulative accuracy scores and total times spent on a test would be expected due to the fact that inaccuracy or error decreases speed as well as cumulative accuracy scores. The equivalence of work limit and time limit scores on low difficulty tests is consequent to the definitions of the former as the average time required to complete an item and of the latter as the number of items completed in unit time. As defined, the measures obviously tend to be related as arithmetic and harmonic means. In some of the more careful studies, characterized by timing each subject on each task, control of accuracy and use of statistical test, speed has been found to be significantly correlated with altitude. In others, it has been found to be unstable, as difficulty of task varies. These more careful studies, in general, have been inconclusive because of extreme skewness and lack of homogeneity of variance of the speed measures or because of loss of independence of the measures when converted into scores based upon group performance on each task, such as percentile ranks and deviation scores.

There is a steadily increasing body of evidence that speed of response, when appropriately measured, is a stable and independent ability. Factor analyses of the intercorrelations of scores on speeded tests of low difficulty have consistently isolated a speed factor. Baxter,[1] Davidson and Carroll[2] and others have found the ability to be elicited by mental test materials of higher difficulty. In the writer's study[5] both a general speed ability and a special speed ability linked to function were found operative in response to four types of mental test materials of varying degrees of difficulty.

The hypothesis of a general ability in speed, independent of mental altitude and independent of function in which it is observed appears to be tenable. The ability has been observed in the testing situation, and, since this situation tends to equalize motivation, the magnitude of the differences observed probably is at a minimum. Is it an important individual difference? At present, little is known about its nature or its correlatives. It may be a native trait. It may be a set for a particular rate of work. A number of factor analysts have concluded that the factor, as elicited by low difficulty tasks, either is an ability or a readiness for discovering and identifying perceptual detail. If so, it may be inherent in the mental act itself, in a habitual way of going about a task, or in efficiency of use of mental power. Perhaps it is a basic component of that psychological complex known as temperament.

Practically, however, these questions are not as important as the questions regarding mental speed correlatives. Do individual differences in speed set off occupational groups or successful workers within groups? Are there antecedent-consequent relations between, say, speed in verbal ability and success in occupations requiring the ready word and pen, such as lecturing and journalism? Is there a speed loading in school marks? Are there racial differences in speed? If speed is a basic component of temperament, does it act as one of the determiners in the subtle process of selecting friends and mates? Is it associated with conservatism-radicalism attitudes? Is it related to persistence? Do individuals differ in variability as well as in mean speed in responding to a set of tasks, and, if so, does the difference in variability provide useful information?

There are a great many provocative problems relating to the rôle of speed in behavior. It is the purpose of this paper to consider some of the difficulties which come up in the measurement of mental speed and to suggest methods of obtaining reliable measures of it.

METHODS AND MATERIALS

Among the kinds of studies of particular interest and importance are (1) the selection of tasks which reliably tap speed, (2) the effect of experimental treatments upon speed, and (3) the correlatives of speed. The first, of course, is basic to any study, and it requires that each individual in the sample under investigation be timed on each task. Individual timing eliminates irrelevant time loss due to search for attractive items and failure to complete undertaken tasks, permits control of the effect of inaccuracy, and provides data for precise study of reliability. Total times on tests in group situations may provide useful information for some purposes, but they do not provide a satisfactory basis for individual comparisons. They tell little about how time is spent and nothing about variability within the individual.

For several reasons typical intelligence test items are well-adapted to studying speed of response. These items tend to require from about eight to forty seconds, on the average, from the individuals for whom they are designed; they are relatively free from familiarity and practice effects; and they can be selected so as to tap reliably a number of partially independent functions. Shorter items are proportionately more affected by errors in timing and time differences ascribable to different rates of reading the stimulus material and indicating the response, while the probability of time loss due to gross irrelevancies tends to be greater in the longer.

There likely are many other sorts of paper-and-pencil and performance tasks which would prove well-adapted to particular problems. For example, extensive sets of arithmetic exercises or monotonous manipulatory tasks might be found appropriate for determining whether variation in speed of response during long and tiring periods of work can be associated with, say, persistence.

It is reasonable to surmise that the tasks should be of completion rather than of multiple choice type, since time pressure likely tends to increase guessing. The relation of speed to motivation in an individual, in general, probably is very close; in fact, if it were possible to obtain a sort of average speed index for an individual, based upon performances under varying degrees of motivation, it might be possible to study motivation quantitatively. Be that as it may, it seems evident that the individual subjects in a speed study should be working for speed and accuracy, under the most comparable motivating conditions possible.

AN APPROPRIATE UNIT OF MEASUREMENT

When a number of short, homogeneous tasks are individually administered and timed in seconds, the distributions of time scores of the individuals in the sample will be characterized by severe nonnormality and heterogeneity of variance. The development of precise statistical tests applicable to such distributions is a matter of great complexity, if possible at all. Thus, the average investigator must either forgo exact statistical *analysis of the speed measures or must find a method of introducing* normality and homogeneity without destroying the independence of the measures.

In his examination of a random sample of the distributions of speed measures of thirty-six high-school students on arithmetic reasoning, number series, sentence completion and spatial relations tests, the writer[5] found the common logarithms of the times in seconds to be distributed normally with acceptably homogeneous variance, as judged by the Chi-square and L_1 criterions. Each test contained about sixty items of mainly twenty to sixty per cent difficulty, in terms of percentages failing. Since that study, the writer has examined a great many sets of response times and has found that taking the common logarithms of the times in seconds has always improved the normality and has usually improved the homogeneity conditions.

It is pertinent here to examine the 'raw' time and log time distributions of several samples of individuals on various kinds of tasks. The distributions in Table I are based on Peak's and

Boring's report[4] of the results of administering Forms A and B of the *Otis Self-Administering Test of Mental Ability, Higher Examination,* to two male and three female university students. The investigators report only the rounded number of seconds for the subjects on the items all worked correctly. (Subject D spent 400 seconds on one of the items. That score is not included in the Table.) Common logarithms with two place mantissas were taken of the times in seconds and multiplied by 100 to

TABLE I.—DISTRIBUTIONS OF CORRECT RESPONSE TIMES IN SECONDS AND LOG SECONDS OF FIVE UNIVERSITY STUDENTS ON OTIS MENTAL TEST-ITEMS

Time in Seconds	Subjects					Time in Log Seconds*	Subjects				
	A	B	C	D	E		A	B	C	D	E
0–4	5	7	4	1	1	26–38	2	1	1		
5–9	36	29	31	13	7	39–51	1	2			
10–14	25	25	24	26	23	52–64	2	4	3	1	1
15–19	14	15	18	21	16	65–77	14	10	10	2	1
20–24	8	10	10	9	17	78–90	13	13	13	6	3
25–29	3	5	1	5	9	91–103	15	14	14	9	10
30–34	3	4	6	7	7	104–116	19	17	18	22	16
35–39		3	1	7	3	117–129	14	16	18	21	17
40–44	2		2	2	2	130–142	10	13	11	13	20
45–49			2	3	4	143–155	4	6	6	8	13
50–54		1		1	3	156–168	2	2	5	11	8
55–59					2	169–181	3	1		3	6
60–64	3			1	1	182–194	1	1	1	2	3
65–69	1					195–207				1	2
70–74				1	1						
75–79		1	1	1	2						
80–84											
85–89											
90–94					1						
95–99				1							
100–104											
105–109					1						
N	100	100	100	99	100	N	100	100	100	99	100
Mean	14.40	14.95	15.50	21.85	25.80	Mean	105.7	107.0	109.5	125.5	131.7
Sum of Squares	15075	12350	13252	25428	35550	Sum of Squares	91935	87152	79733	74020	78755

* Common logarithms of the times in seconds, rounded to two-place mantissas and multiplied by 100.

eliminate decimals. For the sake of table compactness a uniform grouping scheme is employed in forming the distributions. Where rounding to the nearest second could affect the log scores to the extent of changing class membership, the log scores are assigned proportionately to the classes affected. Considering the raw time distributions for Subjects A and E, which appear least similar, the ratios of Fisher's g_1 and g_2 statistics[3] to their standard errors are about 11 and 15 and 8 and 9, respectively. In the log distributions, the ratios are .9 and .3 and .8 and .5. The effect of the log transformations on the variances is very marked. The F-ratios of the raw time variances for Subjects A and E is 2.25; that of the log time variances, 1.17. A ratio of

TABLE II.—DISTRIBUTIONS OF CORRECT RESPONSE TIMES IN SECONDS AND LOG SECONDS OF FOUR HIGH-SCHOOL SENIORS ON MINNESOTA PAPER FORM BOARD ITEMS

Time in Seconds	Subject				Time in Log Seconds*	Subject			
	1	2	3	4		1	2	3	4
0–4	3	1			40–49	1			
5–9	21	19	7	4	50–59	1	1		
10–14	9	11	14	6	60–69	2	1		
15–19	10	4	12	15	70–79	5	5	1	1
20–24	5	2	11	6	80–89	7	4	3	
25–29	4	3	7	7	90–99	7	9	4	4
30–34	1	1	3	2	100–109	8	8	8	3
35–39		1	1	1	110–119	7	4	8	12
40–44		1	1	1	120–129	6	3	8	5
45–49			1	1	130–139	5	2	11	8
50–54		1	1	1	140–149	3	3	8	5
55–59	1				150–159	1	2	3	3
60–64	1				160–169		1	2	2
					170–179	1	1	1	1
					180–189	1		1	

* Common logarithms of the times in seconds, rounded to two-place mantissas and multiplied by 100.

about 1.59 is significant at the one per cent level. It is interesting to note that Peak and Boring, after transforming the times of response by taking "1000 time the ratio of the time of one subject to the total time of the five subjects on any item" observed a difference between the mean times of Subjects A and E of about 11.9 times its standard error. Based upon the log measures, the critical ratio is about 4.3.

Table II gives the distributions of raw and log times of correct response of four college preparatory high-school seniors to the sixty-four items of the Minnesota Paper Form Board, Form A, test. Two of the four subjects were girls. All were of superior academic aptitude and of approximately the same chronological age. The improvement effected by the log transformation is

TABLE III.—POOLED DISTRIBUTIONS OF CORRECT AND INCORRECT RESPONSE TIMES IN SECONDS AND LOG SECONDS OF FIVE HIGH-SCHOOL GIRLS ON THIRTY SENTENCE COMPLETION ITEMS*

Time in Seconds	Correct Response	Incorrect Response	Time in† Log Seconds	Correct Response	Incorrect Response
0–8	17	5	26–38	1	
9–17	36	14	39–51	2	
18–26	18	8	52–64	2	
27–35	13	10	65–77	3	1
36–44	3	5	78–90	7	3
45–53	2	1	91–103	13	3
54–62	1	1	104–116	15	7
63–71	2	3	117–129	17	8
72–80	1		130–142	11	5
81–89			143–155	13	10
90–98			156–168	5	5
99–107		1	169–181	2	3
108–116		2	182–194	2	2
117–125			195–207		3
126–134		2	208–220		3
135–143		1			

* Twenty of the items are reproduced in Table VII.

† Common logarithms of the times in seconds, rounded to two-place mantissas and multiplied by 100.

least observable in the case of Subject 4. For that subject's raw time distribution, the ratios of g_1 and g_2 to their standard errors are 3.4 and 2.5, respectively; for the log time, .2 and .1. Only twenty-eight of the sixty-four items in the test were responded to correctly by all four subjects, and, since the items varied greatly in length, comparisons of means and variances cannot be appropriately made.

The distributions in Table III were obtained from a study of the speed of response of thirty high-school girls to thirty CAVD type sentence completion items.* The means and variances of the log times for five of the girls were approximately equal, and the distributions of the five were pooled. The five were enrolled in other than college preparatory curricula, and their IQ's were 97, 100, 102, 107 and 109. The Table contains four fewer than the possible number of scores, due to failures to respond.

Table IV† gives the pooled raw and log time distributions of correct and incorrect response of ten married pairs to thirty CAVD type sentence completion items. (On eleven occasions items were passed without response, and nine of the times fell between 65 and 134 seconds. The nine were deleted for the sake of table compactness.) In terms of percentages failing, the mid-measure of item difficulty was fifteen per cent; the first- and third-quarter measures were five and thirty-five per cent, respectively. The majority of the variances of the twenty log time distributions were acceptably homogeneous, and the means were distributed normally. The pooled log time distributions in Table IV appear to be the sort expected, under the circumstances, in sampling from normal parents. In the distribution of

---

* In the administration of the sentence completion items, a screening device which could be slipped down the pages of items was used. In determining a subject's speed of response to an item, the watch was started at the instant the screen came to rest, disclosing the item, and stopped at the instant the subject filled in the last blank of that item.

† I am indebted to Miss Mildred Muench of the Clinton, New York, Central School for the data summarized in Table III and to Mr. J. B. Ellery, now at the University of Colorado, for the data in Table IV. The former had investigated the rôle of speed of response in fifteen pairs of close girl friends; and the latter, its rôle in ten young married couples. The former found the between pair speed variance to be greater than the within pair variance at about the five per cent level of significance; the latter observed significance at about the three per cent level.

Table IV.—Pooled Distributions of Correct and Incorrect Response Times in Seconds and Log Seconds of Twenty Adults on Thirty Sentence Completion Items*

Time in Seconds	Correct Response	Incorrect Response	Time in† Log Seconds	Correct Response	Incorrect Response
2–4	56		30–39	7	
5–7	91	9	40–49	7	
8–10	108	10	50–59	18	
11–13	66	10	60–69	31	
14–16	50	13	70–79	42	5
17–19	29	10	80–89	57	5
20–22	20	14	90–99	58	6
23–25	15	9	100–109	77	9
26–28	7	6	110–119	62	14
29–31	7	3	120–129	41	17
32–34	3	8	130–139	33	16
35–37	4	2	140–149	16	12
38–40	4	4	150–159	10	11
41–43	2		160–169	8	8
44–46	4	3	170–179	5	5
47–49	1	2			
50–52	2	2			
53–55	3	3			

* Twenty of the items are reproduced in Table VII.

† Common logarithms of the times in seconds, rounded to two-place mantissas and multiplied by 100.

log times of correct response, the ratios of g_1 and g_2 to their standard errors are .9 and .6, respectively. Twenty of the thirty sentence completion items used in the studies from which the data of Tables III and IV were obtained are reproduced in Table VII.

Not only do log times of response to a number of short, homogeneous tasks tend to be distributed normally within the individual; the log times of a number of individuals on a single task tend to be similarly distributed. The samples of individuals at hand are too small to permit adequate examination of the distributions of log times of response to a single item. When the measures from several items having equal mean lengths and acceptably equal variances are pooled, however, their correspondence to normal form is very evident. The distributions of log and raw times of response of the thirty girls to four such items are shown in Table V. Table VI contains the distributions of the ten married pairs on six items having approximately equal means of log times of correct response. Both tables contain fewer than the possible number of scores due to failures to respond.

TABLE V.—DISTRIBUTIONS OF CORRECT AND INCORRECT RESPONSE TIMES IN SECONDS AND LOG SECONDS OF THIRTY HIGH-SCHOOL GIRLS ON FOUR COMPARABLE SENTENCE COMPLETION ITEMS

Time in Seconds	Correct Response	Incorrect Response	Time in* Log Seconds	Correct Response	Incorrect Response
0–8	2		70–79	1	
9–17	31	7	80–89		
18–26	18	14	90–99	2	1
27–35	10	10	100–109	8	1
36–44	2	9	110–119	12	3
45–53	1	3	120–129	14	6
54–62		1	130–139	10	8
63–71	2	3	140–149	10	8
72–80			150–159	5	9
81–89			160–169	2	7
90–98		2	170–179		3
99–107		1	180–189	2	3
108–116	1	1	190–199		2
			200–209	1	2

* Common logarithms of the times in seconds, rounded to two-place mantissas and multiplied by 100.

The various distributions shown in Table I–VI have been selected as representative of the many possible distributions afforded by the samples of individuals and tasks in hand. They support the hypotheses that (1) times of response to mental test items, measured in log seconds, tend to be normally distributed in the individual and (2) times of response on a single item, measured in log seconds, tend to be normally distributed in a

TABLE VI.—DISTRIBUTIONS OF CORRECT AND INCORRECT RESPONSE TIMES IN SECONDS AND LOG SECONDS OF TWENTY ADULTS ON SIX COMPARABLE SENTENCE COMPLETION ITEMS

Time in Seconds	Correct Response	Incorrect Response	Time in* Log Seconds	Correct Response	Incorrect Response
5–9	6		77–85	2	
10–14	23	5	86–94	3	
15–19	14	6	95–103	7	1
20–24	12	8	104–112	11	2
25–29	5	7	113–121	14	3
30–34	4	5	122–130	8	6
35–39	2	5	131–139	11	7
40–44	3		140–148	6	8
45–49	3	4	149–157	4	7
50–54			158–166	5	3
55–59	1		167–175	2	3
60–64		4	176–184		4
65–69			185–193		1
70–74		1			

* Common logarithms of the times in seconds, rounded to two place mantissas and multiplied by 100.

group of individuals. The distributions of response times to sentence completion items have been utilized only in order to compare different samples of individuals. The writer[5] found speed, measured in log seconds, in arithmetic reasoning and in number series completion to be slightly better distributed with regard to normality and homogeneity and to discriminate between individuals in a sample of thirty-six more sharply than speed in sentence completion. The former two have the additional advantage of being easily scored for accuracy on an all-or-none basis.

The converse of a theorem that the distribution formed by pooling samples from a normal universe is normal would not necessarily hold. But where there are no systematic differences with respect to normality between the samples, it is reasonable to suppose that the variable is distributed normally in the populations sampled.

The hypotheses appear to obtain for a wide variety of tasks from zero to at least about sixty per cent difficulty, defined by percentages failing, and for a wide variety of individuals who are striving for both speed and accuracy in their performance. Performances on tests, however scored, obviously are affected by the individuals and the tasks, but in view of the variety of individuals and tasks here under consideration, it appears *likely that log time will prove a generally useful unit in measuring* speed of response.

CONTROL OF ACCURACY

Given appropriate tasks and units of measurement, the next requirement of a study of mental speed is provision for control of inaccuracy effect, which, the data of Tables III–VII indicate, tends to be substantial.

If two or more individuals respond to the same tasks with the same degree of accuracy, their rates of work can be compared simply and precisely. This ideal situation can be made to obtain by retaining for an analysis only the response times on tasks which all subjects perform correctly. Such experimental control, though the soundest possible, has two disadvantages. First, in practical situations, it is likely to result in enormous loss of data. For example, thirty-six of the sixty-four Minnesota Form Board items would be lost if the data of Table II were limited to items which all four subjects performed correctly. Second, since such tasks are of zero difficulty, they cannot be shown to be measuring anything except speed. Measures based upon them provide no information regarding rate of doing various kinds of work of varying degrees of difficulty.

There are several statistical methods of controlling accuracy effect, at least two of which merit consideration. If approximately equal length tasks, on an average, are available, such that all of the individuals in a sample perform two or more correctly and two or more incorrectly (the more, of course, the better), a precise statistical control is possible. In this situation, mean times of correct and of incorrect response may be computed and classified for conventional analysis of variation. The simplest classification would be a two array table and would permit tests of significance of accuracy and individual effects on mental speed in a particular function. Various extensions of the design suggest themselves. If the tasks can be classified according to two or more levels of difficulty or two or more functions or two or more experimental treatments, a four or more array classification of the speed scores can be made. It will be noted that, provided the log times can be shown to be distributed normally with equal variance, the assumptions underlying variance analysis will be explicitly fulfilled.

A second method of controlling accuracy effect consists in adjusting each response time or speed score for accuracy by regression procedure. In a single task, accuracy being scored '0' or '1,' the residuals from the regression line of speed on accuracy will be merely the difference between the scores and the mean of correct or of incorrect response, as the case may be. There will result as many residual or adjusted scores for each individual as there are tasks; and the mean of these will provide an unbiased estimate of the individual's speed of response free from task and accuracy effects. This method, which the writer has more fully described elsewhere,[5] has the advantage of not obscuring the variability of the original measures, of providing control of possible systematic differences in length of task, and of providing a single measure of individual speed. Because of the latter advantage, the method will be particularly useful in studying mental speed correlatives.

The question of whether speeds of correct and of incorrect response are homogeneous in an individual is an important and interesting one. Theoretical considerations suggest that if a task were unintelligible to an individual, the time of his response would have little meaning. They also suggest that different sorts of inaccuracies or errors will affect response times differently. Empirically, the speeds do appear to possess homogeneity. The evidence indicates consistency in rates of correct and of incorrect response, in general, i.e., individuals tend to maintain their relative positions in rate regardless of accuracy, although there may be significant individual differences here, and, as a speculation, these may be important.

THE RELIABILITY OF SPEED MEASURES

Both of the above methods of statistical control of accuracy make it possible to determine whether the variation in speed of response between individuals is significantly greater than the variation within individuals. The first method, however, since it utilizes the means of correct and incorrect response times, obscures the variability of the original measures. It is possible that significance of differences in speed demonstrated by the method would not obtain if the original measures were given freer play.

The second method provides the data for testing the significance of differences between estimated true speeds of response, with the criterion or error variance, since it is the pooled variances of the adjusted scores within individuals, at a maximum. (It should be noted that the application of the method results in the loss of two degrees of freedom per item.) The test of reliability here is the simple one of determining whether the variation of the means of the adjusted scores is significantly greater than the variation of the scores within individuals, the assumptions being those of normality and of homogeneity of error variance. In his examination of sixteen sets of adjusted speed measures, the writer has found the first assumption generally tenable. The second assumption has generally been in doubt, although not sufficiently to invalidate variance-ratio tests. It appears probable that there are significant and possibly important individual differences in variability in mental speed. The *possibility invites further investigation.*

The second method also makes it possible to estimate the reliability of single items. The correlation of adjusted times on an item with aggregate adjusted times on the remaining items will provide an index of the discriminating power or the reliability of the item as a measure of speed.

The twenty sentence completion items entered in Table VII were common to the studies of mental speed in the three samples of individuals mentioned above. The samples differed in several respects. The median IQ in the sample of thirty-six, consisting of twenty-two boys and fourteen girls, was 116. The boys and girls were enrolled in a college preparatory curriculum and all were of relatively favorable socio-economic background. The median IQ in the sample of thirty girls was 110. The girls were of diverse socio-economic backgrounds and were enrolled in various curricula. The males of the sample of twenty adults, consisting of ten married pairs, were above the 60 percentile for college students on the Henmon-Nelson mental test and were enrolled in a highly selective men's liberal arts college. All were veterans. Nine of the ten wives had two or more years of college training, and all appeared to be of average or better intelligence among college women.

Space does not permit inclusion of the log time and accuracy scores of each individual on each item, but the statistics derived from them are shown in Table VII. A rough estimate of the standard deviations of the log times of correct and of incorrect response, from which the means in the table were derived, indicated that they fell between about 15 and 35. The standard errors of the means, consequently, range from about 2 upward, with the great majority lying between 5 and 10. Reliable differences in both speed and altitude are indicated between the populations of high-school students and the population of adults represented by the samples. The differences between the populations of high-school students approach significance on few items, although the sample of thirty has a mean cumulative accuracy score on the twenty items of 12.2 against a mean of 13.7 for the sample of thirty-six. The difference is about 2.5 times its standard error.

The extent of correspondence of the ways the items worked in the three samples is interesting. Considering the samples of thirty and thirty-six, the rank-difference correlation coefficients between mean log times of correct response, between mean log times of incorrect response, and between item difficulties are .86, .75, and .76, respectively. For the samples of thirty and twenty, the coefficients between mean log times of correct response and between item difficulties are .85 and .72. For the samples of thirty-six and twenty, the coefficients between mean log times of correct response and between item difficulties are .93 and .80.

The item reliability coefficients shown in Table VII were determined by rank-difference correlation of adjusted times on each item with total adjusted times on the twenty items. The majority of the coefficients in the two samples of high-school students are significant. The differences between coefficients in the two samples are such as would be expected in sampling. Only item 15 raises doubt. There, taking into account the loss of two degrees of freedom, the null hypothesis is discredited at about the three per cent level. In the sample of adults fewer

TABLE VII.—MEAN LOG TIMES OF CORRECT AND INCORRECT RESPONSES, DIFFICULTY AND RELIABILITY OF EACH OF TWENTY SENTENCE COMPLETION ITEMS FOR THREE SAMPLES OF INDIVIDUALS

Sentence	Sample of Thirty High-school Girls				Sample of Thirty-six High-school Girls and Boys				Sample of Twenty Young Adults			
	Correct Response	Incorrect Response	Per Cent Failing	Reliability	Correct Response	Incorrect Response	Per Cent Failing	Reliability	Correct Response	Incorrect Response	Per Cent Failing	Reliability
1. Two pounds of silver are ____ more than two pounds of iron.	81	128	33	.36	86	100	33	.55	71	104	10	.14
2. It ____ strength to ____ a heavy weight.	97	115	30	.12	100	131	39	.45	79	118	20	.09
*3. We are bound to ____ if every one does his ____.	98	117	40	.47	99	102	14	.65	83	94	5	.08
*4. Rich and ____ alike are seldom happy without good ____.	109	132	10	.36	101	116	17	.33	83	—	0	.30
5. A body of ____ entirely surrounded by ____ is called a ____.	110	142	27	.39	114	148	22	.38	103	117	15	.63
6. It may ____ effort and a long ____, but the result is sure.	112	119	27	.62	112	138	36	.38	90	114	30	.54
*7. A ____ is judged by the ____ he ____.	115	155	7	.52	108	128	11	.62	87	—	0	.50
8. There is no ____ on earth ____ cannot bear ____ misfortune.	120	145	33	.39	114	126	11	.56	108	—	0	.62
*9. Time ____ ____ worth more ____ money.	123	161	40	.37	138	161	36	.34	122	136	25	.33

*10. You may ____ some of the questions even though you are not certain your ____ are ____.	125	123	23	.65	129	143	31	.61	111	—	0	.48
*11. Poverty cannot ____ a man ____ is intelligent and who ____ hard.	125	150	20	.49	122	149	11	.62	119	178	10	.59
12. Many new ____ are printed every year, but some wise ____ prefer to ____ the old ones.	128	140	20	.65	128	136	22	.75	105	167	5	.63
13. The leader was unjustly ____ by those who, in his place, would have probably ____ no differently.	130	141	47	.62	121	137	39	.63	99	105	5	.06
14. Most men ____ themselves ____ more kindly than their ____ judge them.	130	152	50	.43	143	140	39	.29	126	119	35	.56
15. She left a ____ for the people whom she thought would ____ during her ____.	134	161	40	.72	140	148	53	.31	120	155	25	.24
16. ____ spite ____ many severe ____ he is still alive ____ the ____ of ninety-one.	135	174	43	.51	137	152	28	.73	125	150	10	.09
17. He will do as you request ____ ____ his own feelings.	137	137	90	.41	138	135	42	.39	122	131	35	.42
18. As ____ the treasure he had come to seek, probably it existed ____ in his own ____.	141	175	67	.54	135	168	64	.47	109	143	45	.13
19. No ____ is powerful ____ to ____ two and two be five.	142	163	50	.27	130	153	42	.38	131	153	55	.65
20. If ____ the ____ of the year were holidays, playing ____ be ____ tedious ____ working.	154	176	63	.60	156	171	44	.45	143	157	35	.41

---

* Only sentences 3, 4, 7, 9, 10, and 11 were constructed for the three studies. The other sentences were taken from Thorndike's lists.[6]

than half of the coefficients appear to be significant, and several coefficients indicate significant differences between the population of adults and the populations of high-school students. All in all, however, the data of Table VII indicate that item times are at least as stable as item difficulties. In sentence completion, at least, the possibility of developing quite reliable speed measures and norms appears promising.

SUMMARY

In this paper several questions relating to the rôle of mental speed in behavior have been raised, and a method of obtaining normally distributed measures of speed has been proposed. Considerable evidence has been presented in support of the hypotheses that (1) times of response to mental test items, measured in log seconds, tend to be normally distributed in the individual and (2) times of response on a single item, measured in log seconds, tends to be normally distributed in a group of individuals. Two statistical methods of controlling accuracy have been suggested. The extent of the stability and consistency of speed of response to sentence completion items for three samples of individuals has been briefly described.

REFERENCES

1) Brent Baxter, "An Experimental Analysis of the Contributions of Speed and Level in an Intelligence Test," *Journal of Educational Psychology,* 32:285–96, 1941.

2) W. M. Davidson and J. B. Carroll, "Speed and Level Components in Time-limit Scores: A Factor Analysis," *Educational and Psychological Measurement,* 5:411–27, 1945.

3) R. A. Fisher, *Statistical Methods for Research Workers,* London: Oliver and Boyd, 8th Edition, 1941, pp. 50–52.

4) Helen Peak and E. G. Boring, "The Factor of Speed in Intelligence," *Journal of Experimental Psychology,* 9:71–94, 1926.

5) Merle W. Tate, *Individual Differences in Speed of Response in Mental Test Materials of Varying Degrees of Difficulty,* Unpublished Doctor's Dissertation, Cambridge: Harvard University, 1947. (See Vol. 8, pp. 353–374, *Educational and Psychological Measurement,* for a digest of the dissertation.)

6) E. L. Thorndike, *Measurement of Intelligence,* New York: Bureau of Publications, Teachers College, Columbia University, 1926.

GAINS IN SCHOLASTIC APTITUDE UNDER HIGHLY MOTIVATED CONDITIONS

GEORGE E. SCHLESSER
Department of Education, Colgate University

Results of an experimental sixteen-week intensive training program for adult servicemen showing change in scholastic aptitude seem worth while adding to the increasing amount of evidence concerning the constancy and change of scholastic intelligence. At present the majority of studies have used data on students of school and college age, but have not found adult groups as accessible for retesting. The study has bearing on two issues in educational psychology. (1) Does scholastic intelligence as measured by tests stop growing at age sixteen, in the early twenties, or later? (2) How great a change would highly motivated schooling make in the scholastic intelligence of young adults?

On the latter issue, data on mountain, gypsy, and canal boat children, who are educationally isolated, have revealed a drop of over twenty-five IQ points or about one hundred sixty scaled score points* in an eight- or nine-year period. Soldiers of World War I from educationally backward states had a lower IQ than did soldiers from states with advanced educational systems. Children adopted into favorable foster homes also show an average increase of about five IQ points. These facts have led Bingham to conclude that "there can be scarcely any doubt today that schooling and other educative experiences tend to increase a child's intelligence." Can adults improve their scholastic intelligence under favorable conditions even after they are twenty years old?

In this study, scholastic intelligence is measured by the ACE Psychological examination and high motivation took place during Navy Pre-Midshipmen Refresher Program at Colgate University. The purpose of the program was to give enlisted men with an average of three years' college training an opportunity of completing midshipmen's training on an equitable basis with other officers candidates who had completed college work. Such men were recommended as candidates for this V-7 program by the captains of their ships or by their shore-based commanding officers. Every one wanted to become an officer and knew that by completing the refresher program and by passing his sixteen weeks' training in Midshipmen's School he definitely would be made an officer. Here an opportunity for up-grading was given in which both the goal and the steps to reach it were defined. It was Colgate's task to see that each man's aptitudes and achievements were so appraised that he could be given training at the proper speed and level. Some men would need sixteen weeks of refresher work; some, eight weeks; and few would be able to go on immediately. The refresher course consisted of English, mathematics, engineering drawing and physics and physical training. In an eight or sixteen weeks' course, a man had to take a full year's work in physics, mathematics, algebra, trigonometry, and navigation and a full year's work in English and an equivalent of a full year's work in engineering drawing. Physical training and drill were added, so that all the waking hours of a man's time on week days were completely consumed.

---

* In this paper one hundred scaled score points equal 1.00 SD and sixteen IQ points equal 1.00 SD.

Evidence from two sets of interviews indicates that the drive to succeed was very strong. A new group entered each month starting in July, 1944, through September, 1945. At the time of entrance each man was given the 1941 Edition of the ACE Psychological Examination and other tests. Those who were to remain in the program sixteen weeks, actually seventy-five per cent of the total, were given the 1942 Edition of the Psychological Examination after twelve weeks. These editions according to the norms were nearly comparable and were constructed as equivalent forms. The norms show that the 1942 edition actually is four points or seven percentiles more difficult, so that actual change is greater than apparent change in raw score.

A description of the nature of the population will be important for later interpretation. The men were 24.5 years of age, had been out of college about 3.2 years after going to college at least two and an average of approximately three years. Fifty-two per cent were men with overseas duty. Most of the remainder had transferred from an Air Cadet (V-5) program. They came from all forty-eight states with the largest portion, ten per cent, having attended high school in New York State. Next ranking states were California, Illinois, Pennsylvania and Texas, respectively. A sample was slightly below average in scholastic intelligence for college freshmen (40th percentile) at the time of initial entrance to college six years earlier. [Data on thirty-six men for whom comparable earlier tests were available are presented in the third succeeding paragraph.]

For purposes of this study, five classes from September, 1944, to January, 1945, inclusive were selected to determine whether any unusual gain in scholastic aptitude was to be found. Table I shows that for these two hundred sixty-one men the average gain was from the 56th percentile to the 78th percentile, or a gain of twenty-two percentiles. Every one of the five groups gained more than twenty percentiles. At the time of entrance the average score for the groups was 110.7 and at the close of the twelve-week period the score was 125.7. This difference is over seven times what might be expected from chance (CR = 7.1).

TABLE. I.—PRETEST AND RETEST RESULTS ON THE ACE PSYCHOLOGICAL EXAMINATION IN A TWELVE-WEEK PERIOD

Group	N	Pretest Average	Retest Average	Raw Score Gain	Pretest SD	Retest SD
Total	261	110.66	125.70	15.04	22.98	22.08
A	40	104.90	122.60	17.70	21.18	21.00
B	54	112.40	127.04	14.64	19.08	19.56
C	53	110.84	125.54	14.70	23.28	23.88
D	51	109.04	122.54	13.50	20.94	18.96
E	63	114.08	129.20	15.12	24.62	23.82

At the same time the men remained in relatively the same position with respect to each other. This is indicated by a correlation of .88 between the pretest scores and the retest scores of the two hundred sixty-one men—a coefficient which is almost as high as the reliability of the test, even though it includes a twelve-week lapse of time.

It was thought possible that the gain might be due to loss of scholastic aptitude during their several years of service with the fleet and that this increase represented a return to their normal score. To test this, letters were sent to their earlier colleges for the classes entering in September, 1944, and November, 1944, to obtain scholastic aptitude scores given at the time of entrance to college. Thirty-six of these returns were such that results could be directly compared to the later pretest results. The raw score mean for the thirty-six men at the time of college entrance was 94.9, their mean at the time of pretest was 108.4, and their mean at the time of retest was 122.2. Thus, over a period of about six years they had gained thirteen and one-half raw score points or sixty scaled score points. Whether this is lower than normal growth cannot be answered, but there is some evidence to show that greater mental growth should be expected during the four college years.[1] The sample appears to be about two or three points lower in scholastic intelligence than the large group.

There are several possible factors that may account for some of the gain of the large group: (1) practice effect, (2) regression if below average individuals are used, (3) normal growth during the twelve weeks, (4) genuine growth in mental alertness during the refresher program. The practice effect of retaking CEEB scholastic verbal aptitude test is a gain of about twenty scaled score points, even though different forms are used.[2] It seems reasonable to assume, then, that the effect of such practice on the ACE Psychological Examination accounts for not more than about twenty of the sixty-eight scaled score points gained by the V-7 men.

The amount of change due to regression, the second factor, should be negligible, for the sample indicated that the large group averaged between the 40th and 45th percentile according to national norms at the time they entered college. Likewise the change due to normal growth, factor three, during the three-month period should be very small. One study showed the growth to be about ten scaled points for three months for seventeen and eighteen year olds.[2] Normal growth at the average age of twenty-five is probably less than this. If ten scaled points are assigned to the influences of both the second and third factors a genuine growth of thirty-eight scaled score points or a gain of eleven percentiles may be attributed to the special educational experience.

---

[1] T. M. Livesay, "Does Test Intelligence Increase at the College Level," *Journal of Educational Psychology*, Volume 30, 1939, pp. 63–68.

[2] Richardson, Pearson, "Effects of Growth and Retesting on SAT Verbal Scores," *The College Board Review*, Vol. I, No. 5, Fall, 1948, pp. 57–71.

The question of shift in rank in scholastic aptitude also has significance for guidance. In the sample group of this study the correlation between earlier entering college scores and pretest score was .60. (The correlation between pretest and retest has already been reported to be .88). Another study has shown the correlation during a nine-year adolescent period to be about .50.[3] These two correlation coefficients are in relative agreement. If substantiated by other studies, they suggest that there may be considerable shifting in the relative rank of young adults in scholastic aptitude. Test scores of entering freshmen may be only moderately predictive of senior or adult scholastic intelligence.

CONCLUSIONS

1) An average gain of twenty-two percentiles (sixty-eight scaled score points) in scholastic aptitude over a twelve-week period was consistently shown by five classes in which the average age was 24.5 years.

2) About half of this gain might be accounted for by practice effect, regression and maturation, but a significant average gain of at least eleven percentiles or thirty-eight scaled score points still remained which may be attributed to the effects of the training program.

IMPLICATIONS

1) Counselors of freshmen can encourage freshmen who rank below average in scholastic intelligence. Under the right motivation and working conditions whole groups can increase their tested scholastic intelligence by twenty-two percentiles. Many individuals can do better; some may do worse.

2) The entering freshman's scholastic intelligence percentile seems to remain very stable in a relative sense over a three month's period, but it may fluctuate over a period of years sufficiently to make the retesting desirable for graduate schools or for industries that use scholastic intelligence as one of several selection devices.

---

[3] H. E. Jones, and H. S. Conrad, "Mental Development in Adolescence," *43rd Yearbook,* The National Society for the Study of Education, Part I, p. 169.

COMMENTS

Interpretation in this study is admittedly conservative. The practice effect on the ACE Psychological Examination may be less than the twenty scaled points of the SAT Verbal Examination, because of the practice exercises in the former. Also if the four raw score or eighteen scale score points due to the greater difficulty of the 1942 edition were added to the apparent gain, the change during the twelve-week intensive educational program would be even more impressive.

INTERGROUP ATTITUDES OF GENTILE, JEWISH, AND APACHE INDIAN CHILDREN

ROSE ZELIGS
Avondale Public School, Cincinnati

The purpose of this study is to present the intergroup attitudes of Gentile, Jewish, Apache Indian, Seventh Day Adventist, delinquent, rural, and city children, as expressed on the Zeligs Intergroup Attitudes Test.

THE SUBJECTS

The subjects of this study were one hundred sixty-three Jewish, thirty Protestant, and seven Negro twelve-year-old children studied in 1931; who attended the AV suburban public school in Cincinnati, Ohio; one hundred fifty Jewish, thirty-three Protestant, and nine Negro twelve-year-old children who attended the same school in 1935; one hundred fifty Jewish, twenty Protestant, and four Negro twelve-year-old children who attended the same school in 1944; forty-two Protestant twelve-year-old children studied in 1935 who attended the PR public suburban school in Cincinnati, Ohio; one hundred thirty-seven Apache Indian children ranging from thirteen to seventeen years in age, from Whiteriver, Arizona, studied in 1934; fifty-nine Cincinnati high-school students; forty-six rural northern Ohio high-school students; twenty-three delinquent girls; and eleven Seventh Day Adventist children, studied in 1944.

TECHNIQUES AND PROCEDURES

The children were given the Zeligs Intergroup Attitudes Test which requires the subjects to indicate their willingness to accept members of each of thirty-nine races and nationalities for the relationships of cousin, chum, roommate, playmate, neighbor, classmate, and schoolmate, by writing 'yes' for the relationship and to express their unwillingness by writing 'no' for the relationship indicated. The index of friendliness was computed by adding the 'yes's' expressed by all the subjects towards each race and nationality and computing its percentage of the total possible number of 'yes's.' The total number of 'yes's' on the entire test was used as the child's score in intergroup friendliness.

THE FINDINGS

In 1931 the Gentile children of the AV. School scored somewhat higher on the Intergroup Attitudes Test than the Jewish children, having an average score of 124 as against 117. In 1935 the Jewish children scored 125 as against 107 for the Gentile children of the AV. School and 134 for the Gentile children of the PR. school. The average of 121 for all the Gentile children studied in 1935 indicates little difference between the Gentile and Jewish children in tolerance. The average score for the boys these various times was always somewhat higher than for the girls. The Apache American Indian children of Whiteriver, Arizona, studied in 1934, whose average was 96, scored significantly lower than the twelve-year-old Cincinnati children.

The social distance the children expressed for the different relationships shows that 'cousin' and 'roommate' are considered the most intimate relationships by all the groups studied. In 1931 the Jewish children were more particular than the Gentile children for those relationships, but in 1935 the Gentiles were more particular. In 1935 the difference between 'cousin' and 'chum' was the same for both groups of the AV. School. The PR. School children were slightly more particular about 'roommate' than 'cousin.' The difference between 'cousin' and 'chum' was less for these children than for those of the AV. School. The Jewish children tend to make more definite differentiation than the Gentile children between the most distant and least distant relationship. The Indian children expressed less difference than the white children between these relationships.

Table 1 gives the degree of friendliness toward thirty-nine races and nationalities expressed by the twelve-year-old children in 1931 and 1935, and by the Indian children in 1934. All groups except the Indian children ranked American first. The Indian children ranked American second. In 1931 the Jewish children ranked German Jew second. American, English, Dutch, Irish, American Indian, Canadian, French, and German, were ranked high by most of the groups, while Finn, Roumanian, Hindu, Bulgarian, Mulatto, Bohemian, Syrian, Negro, Servian, and Armenian were ranked low. The Jewish children tested in 1931 showed slightly less favor than those tested in 1935 towards American Indian, French-Canadian, Italian, Mexican, Czechoslovakian, and Finn. The 1935 Jewish subjects showed greater

Table 1.—Index of Friendliness toward Races and Nationalities Expressed by Twelve-year-old Jewish and Gentile Children and by Apache American Indian Children. [Expressed in Percentage.]

	1931		1935		1935 PR. School Gentile	1935 AV. & PR. School Gentile	1934 Apache Indian
	AV. School		Jew	Gentile			
	Jew	Gentile					
American	96	99	97	100	100	100	85
English	81	89	84	84	95	90	65
Irish	53	70	53	61	77	70	48
Dutch	68	72	64	57	76	68	65
American Indian	50	61	57	60	74	67	89
Canadian	69	64	61	40	73	58	29
French	72	85	71	60	71	66	42
Scotch	45	58	49	43	70	58	38
French-Canadian	44	50	52	35	66	52	33
Swedish	51	60	50	43	65	55	30
German	63	77	36	51	63	57	36
Italian	41	47	50	48	62	56	41
Dane	36	37	40	40	60	49	17
Japanese	51	62	48	36	60	49	46
Norwegian	49	56	48	42	59	51	22
Polish	54	34	54	33	58	47	25
Mexican	42	51	50	34	57	47	52
Spanish	48	56	52	49	55	52	45
Filipino	35	52	42	23	53	40	33
German Jew	85	61	77	39	46	43	36
Russian	58	41	59	37	45	42	38
Greek	44	48	47	45	45	45	32
Turk	33	36	39	28	44	37	37
Chinese	36	39	42	32	44	30	35
Portuguese	29	28	39	18	44	32	27
Arab	24	32	30	25	44	35	24
Hungarian	37	40	40	31	41	37	26
Czechoslovakian	30	35	42	29	41	37	28
Russian Jew	80	50	77	37	39	38	27
Finn	17	34	35	19	37	29	28
Roumanian	38	24	40	21	34	28	18
Hindu	22	21	26	16	32	25	17
Bulgarian	31	29	37	30	31	30	25
Mulatto	10	9	10	5	26	16	24
Bohemian	13	15	20	12	20	16	18
Syrian	12	8	19	14	20	17	18
Negro	18	20	21	13	15	14	28
Servian	11	11	17	12	12	12	18
Armenian	11	10	11	11	11	12	26

antipathy towards the Germans, assigning them twenty-ninth place as against the eighth place given by the group tested earlier. The Gentile children of the AV. groups, tested in 1931, showed greater favor than the 1935 group towards the Canadians, Germans, Japanese, Mexicans, Filipino, and German Jew, and less favor towards the Spanish.

A comparison of the 1935 Gentiles of the AV. School with those of the PR. School indicates more favorable attitudes by the AV. subjects towards the French, Germans, and Italians, and greater antipathy towards the Scotch, Canadians, French-Canadians, Japanese, Polish, Mexican, Filipino, and Portuguese. The Gentile children of the AV. School, who came in contact with many Jewish children, show more favor towards German and Russian Jews than the Gentile children of the PR. School where no Jewish children were enrolled.

Comparing the Jewish with the Gentile children of both schools, the Jewish children show much more favor towards German Jews and Russian Jews, Russian, and Polish, while the Gentiles favor the Irish, Scotch, German, and American Indian. In General, these reactions are in keeping with the ethnic background of the children.

Comparing the Indian with the Gentile children, differences represent more favorable attitudes of the Indian children towards the Japanese, Mexicans, Spanish, German Jew, Turk, Chinese, and Negro, and less favorable attitudes towards the French-Canadian, Swedish, German, Dane, Norwegian, and Polish. Many of these differences may be explained by the ethnic backgrounds and contacts of the children. There was close agreement between the two groups in their attitudes toward the other twenty-six races and nationalities on the test.

There seems to be a general pattern among all the groups. This pattern tends to favor the North European peoples with special ranking for the English, Irish, Dutch, Canadian, French, Scotch, French-Canadian, Swedish, German, and Italian. This cultural pattern of group prejudice is similar to that of college students studied by Thurston[2] and of adults studied by Bogardus.[1]

In 1944 the test was given to city high-school students, rural high-school students, delinquent girls, and Seventh Day Adventist adolescent children. The city high-school students, with an average score of 169, were the highest. Then came the sixth-grade children with an average of 160, the Seventh Day Adventists with an average score of 151, and the rural high-school students with an average of 144. The delinquent girls with an average score of 117 were much lower than the other four groups. It is possible that this low score is due to lack of knowledge about some of the races and nationalities rather than to unfavorable attitudes. These girls were low in intelligence, their intelligence quotients being between 80 and 90.

The social distance expressed for the different relationships showed that all five groups ranked cousin first, roommate second, and chum third. The sixth-grade and the city high-school students were more particular about neighbor than were the rural high-school students, the Seventh Day Adventists, and the delinquent girls. The difference between the most intimate and least intimate relationship was forty per cent for the sixth grade; thirty-eight per cent for the city high-school students; thirty-five per cent for the rural high-school students; forty-six per cent for the Seventh Day Adventists, and fifteen per cent for the delinquent girls.

The ranking of the thirty-nine races and nationalities were noted. All five groups ranked American first. The races and nationalities towards which the highest degree of friendliness were expressed in 1944 were American, English, Canadian, Hawaiian, South American, Irish, French, Australian, American Indian, and Dutch. At the bottom of the list, showing the least friendliness expressed towards them, were German, Dane, Bulgarian, Negro, Roumanian, Portuguese, Arab, Finn, Japanese, Hindu, Albanian, and Mulatto.

There was close agreement among all five groups in their attitudes toward American, English, Dutch, Norwegian, Roumanian, and Albanian. Four of the five groups showed close agreement in their attitudes toward Canadian, South American, Hawaiian, Chinese, Irish, Australian, Swiss, Spanish, Greek, Polish, French-Canadian, Italian, Czechoslovakian, Hungarian, Bulgarian, Negro, Portuguese, Arab, Finn, Japanese, Hindu, and Mulatto.

The Seventh Day Adventists were more favorable toward the Jews than were the high-school students or the delinquent girls. Some of the differences may be explained by the racial and national background of the subjects. Except that the present subjects rank German and Japanese much lower, and Russian and Chinese higher than in earlier studies,[3] the general pattern of the five groups is similar to the attitudes of children studied before 1944. The large number of Jewish children in the sixth grade of the AV. School and of Negro children among the delinquent girls studied explains the high ranks given the Jew and Negro by those subjects. In most other cases there is much agreement among the five groups in their attitudes.

SUMMARY

The trend in children's attitudes is definitely toward better understanding. With the exception that children show more favor towards their own racial, religious, or national group, the general pattern of likes and dislikes expressed by the Jewish, Gentile, and Indian children is very similar. Gentile children who came in contact with Jewish children showed more favor towards Jews than those who had few contacts with them. The war undoubtedly made children more conscious of their attitudes. Where formerly the tendency was to favor those who were like them in color and culture, such as the North Europeans, the trend changed towards favoring allies, regardless of other factors. Although the children do feel closer to people who are similar to Americans in color, culture, and language, the trend is towards a wider outlook. The Germans were disliked while the Chinese and Filipinos were admired as friends and allies. There was little difference in the scores of the sixth-grade and high-school students. The rural children scored somewhat lower than the city children. The children studied in 1944 were less afraid of differences in race or culture than the children studied earlier.

REFERENCES

1) E. S. Bogardus, *Immigration and Race Attitudes.* New York: D. C. Heath and Co., 1928, p. 268.

2) L. L. Thurstone, "An Experimental Study of Nationality Preferences." *Journal of General Psychology*, 1928, 1, 405–425.

3) Rose Zeligs, "Children's intergroup attitudes." *The Journal of Genetic Psychology*, 1948, 72, 101–110.

BOOK REVIEWS

PALMER O. JOHNSON. *Statistical Methods in Research.* New York: Prentice-Hall, Inc., 1949, pp. 377.

With the increasing use of statistical techniques in research, greater emphasis must necessarily be placed on the interpretative function of statistics. The statistician needs to know more than how to collect and describe large masses of data. In recognition of this need, Dr. Johnson has prepared a book which stresses understanding of the theoretical bases and the underlying assumptions of methods used by the research worker. His stated aim is to supply students with a book built on the recent advances in theory and practice pertaining to the problems of testing statistical hypotheses and of statistical estimation.

Those who are particularly interested in psychology and in education will be pleased to note in the preface that "the book . . . was developed . . . primarily for graduate students in education and in psychology." It is also reassuring to find on page 14 the statement that the book "is written for readers without specialized mathematical training." But despite these statements and the fact that *Statistical Methods in Research* is a scholarly work which includes a considerable number of illustrative applications of principles, it is generally not suitable as a text for graduate students in psychology and education.

The book begins with a discussion of probability and likelihood. This is followed by sampling distributions and the building up of statistical models against which experimental results may be checked, e.g., the normal distribution, t, and chi-square. A short and heavy discussion of the general theory of testing statistical hypotheses leads into a long chapter devoted to current procedures. Illustrative problems demonstrate the use of various tests of significance. Chapter VI is entitled "The Estimation of Population Parameters." This includes a concentrated theoretical treatment of estimation (point estimation and estimation by interval, fiducial limits and confidence intervals) and the solution of a number of problems of interval estimation. This is followed by a discussion of normal and normalized distributions in which Pearson's and Fisher's methods for testing the normality of a sample are demonstrated. The succeeding chapter, given over to the statistical analysis of data under non-normal assumptions, should be of particular interest to those concerned with data which consist of rankings. Chapter IX is a very good discussion of the various kinds of sampling systems and the types of errors that may be encountered in investigation by sample. Chapters X and XI deal with analysis of variance and covariance and the application of these techniques in practical problems. Chapter XII, on the principles of experimentation, is a clear discussion of the self-contained experiment and of the rôle of statistics in the design of experiments. Chapter XIII illustrates the application of principles of experimentation in various kinds of design, and the last chapter carries through the complete analysis of a regression problem and a problem in the use of discriminant function.

Although the major topics are presented in logical order, some questions may be raised about the placement and emphasis of sub-topics. Yates' correction for continuity in connection with chi-square is noted very briefly and the reader is referred to a reference. Since this correction is frequently called for in psychological experiments, it is unfortunate that more space is not devoted to it. On the other hand, three tests of homogeneity of variability are described and illustrated when one method would seem sufficient. Degrees of freedom are mentioned a number of times in the first five chapters but the concept is not explained until the close of Chapter VI. The analysis of covariance is briefly defined and the problem illustrating its application is much too complex to clarify the use of the method. But these criticisms are relatively minor when compared with the general comment that the book is extremely difficult reading. Psychology and education students who have had no more than one year or one course in descriptive statistics will generally find that they are beyond their depth throughout large portions of the text. Calculus is rather infrequently used but there is great stress on the mathematical formulation and development of statistical concepts. The style in many places is similar to that of highly technical journals. In a book of reasonable size, this style is perhaps unavoidable when the presentation has been based on original and secondary sources of mathematical statistics. But it limits the use of the book to those readers who have a solid foundation in mathematics or to those who have had considerable prior statistical training.

It is the reviewer's opinion that *Statistical Methods in Research* will be found most useful as a reference text or as a course text for those with a major interest in statistics. For most graduate students of psychology and education the book would be of very limited value. JEROME E. DOPPELT

The Psychological Corporation

PAUL F. LAZARSFELD, AND FRANK N. STANTON, Editors. *Communications Research* 1948–1949. New York: Harper and Brothers, 1949. pp. 332.

The field of communications research differs in emphasis from educational research in being concerned more with the impact of mass media, such as newspapers and radio, on persons of all ages, especially adults, usually in out-of-school situations. These differences in substance are not intellectually significant, however, and readers of the JOURNAL will find much in this volume that is interesting and provocative.

The eight studies here collected fall into three methodological categories. The first, intensive interview studies, consists of (1) "The Children Talk about Comics" by Katherine M. Wolf and Marjorie Fisk; (2) "Research For Action" (a study of radio's morning audience) by Paul F. Lazarsfeld and Helen Dinerman; (3) "What 'Missing the Newspaper' Means" by Bernard Berelson; (4) "The Analysis of Deviant Cases in Communications Research" by Patricia L. Kendall and Katherine M. Wolf; and (5) "Patterns of Influence: A Study of Interpersonal Influence and of Communications Behavior in a Local Community" by Robert K. Merton. What these studies have in common is a highly insightful, well-nigh clinical, analysis of responses obtained by skilled interviewers from samples of persons chosen because of their special importance to the problem under attack. These studies are rich in *verstehen;* the designs, analyses and interpretations reflect intimate acquaintance with the research area, immersion in the data, and creative imagination in synthesizing hypotheses and conclusions.

For lack of space, we shall exemplify this group of researches only with the study of comics. We are given classifications (1) of comics into 'funny animal,' 'adventure' and 'educational' types, (2) of comic reading into three chronological stages, (3) of functions of comics into the *Alice in Wonderland,* or variety and action function, the 'batman' function (the invincibility of the vincible); and the *Reader's Digest* function, (4) of readers into fans, moderates, and indifferents or hostiles. Each of these classificatory schemes is exhibited in relation to various others. Psychological ideas are used freely, among these being 'ego-strengthening,' 'projection,' 'identification,' and 'adjustment.'

The recurrence of trichotomies makes us wonder, of course, whether they are intrinsic to the data or merely a Procrustean methodological rut. But the qualitative descriptions of these conceptual schemes and the anecdotal quotations illustrating them do succeed in bringing the discussion to life in a manner seldom seen in research reporting. We hope it was not a necessary concomitant of this liveliness that few of the tables showing relationships are summarized in coefficients of any kind, and no tests of statistical significance are reported. So important a conclusion as the following is buttressed by a single sentence devoid of either a coefficient or a level of confidence: "Cross classification of children by degree of psychological adjustment and degree of interest in comic reading reveals that only one fan is psychotic and a bare handful of fans are normal children. But a marked correlation is revealed between neuroticism and being a fan. Of all the neurotic children, fifty per cent are fans and only thirty-three per cent moderate readers, while for problem children the proportion is exactly reversed, fifty-four per cent being moderate readers and only thirty-two per cent fans." (p. 29)

Somehow we missed these traditional impedimenta of educational research. Perhaps they are absent for the sake of readability. This cavalier approach to statistical description and inference prevails throughout the book. This is hard to understand since other evidences of research integrity and skill are so abundant. Furthermore, in several other places, these papers show much less concern for journalistic vigor. We wish an explanation had been made of this disregard for what so many graduate students have had to learn about statistical methods in research.

The second category of papers, statistical reports, consists of (1) "An Analysis of Radio's Programming" by Kenneth Baker; and (2) "Overlapping Magazine Reading: A New Method of

Determining the Cultural Levels of Magazines" by Babette Kass. The former is a straightforward sampling survey of facts about the programs broadcast by commercial stations during a typical week in 1946. Apart from the facts themselves, this study is noteworthy primarily for the thoroughness and sophistication evident in the description of procedure.

The study of magazine reading is devoted to the development of a coefficient, called 'phi,' of overlapping reading. Phi equals $c/\sqrt{ab}$ where a is the number reading Magazine A, b is the number reading Magazine B, c is the number reading both A and B. Its correlation with impressionistic judgment rankings of magazines is high, the deviations from perfect correlation are analyzed, and limitations of the method are described. Since this phi coefficient correlates so highly with the more easily obtained ratings of magazines by judges, it will probably not be used in practice. Rather than validating phi against impressionistic judgments, we would set up *phi as the criterion.*

The last study is "Domestic Broadcasting in the U.S.S.R." by Alex Inkeles. Based largely on material in Soviet radio 'trade' journals, this more sociological and politico-economic kind of communications research can furnish perspective for educational and social psychologists interested in comparing relationships between social systems and cultural forces.

All in all, the volume combines extraordinary virtues with defects we might never have noticed had not the general level been so high. Hence, for a change, we find ourselves feeling needs for less facile hypothesizing, for greater integration of typologies and constructs from one study to the next, and for more statistical timidity in drawing conclusions. It is also fair to say that these communications researchers would find much to emulate in the superior attention to previous work and the more adequate presentation of statistical evidence that characterize most research in educational psychology. But as models of research approaches that can *fruitfully* be *transferred* to educational problems more traditionally conceived, these studies deserve careful reading by students of educational psychology.

N. L. GAGE

University of Illinois

PUBLICATIONS RECEIVED

Anne Anastasi and John P. Foley, Jr. *Differential Psychology.* Revised Edition. New York: The Macmillan Co., 1949, pp. 894. $5.50.

Ralph F. Birdie, Director. *The Counseling and Guidance Use of Test Scores.* Minneapolis: University of Minnesota Press, 1949, pp. 34. (paper) $1.00.

Homer L. J. Carter and Dorothy J. McGinnis. *Reading Manual and Workbook.* New York: Prentice-Hall, Inc., 1949, pp. 120. (paper)

Alphones Chapanis, Wendell R. Garner, and Clifford T. Morgan. *Applied Experimental Psychology: Human Factors in Engineering Design.* New York: John Wiley and Sons, Inc., 1949, pp. 434. $4.50.

Amy Bishop Chapin and Ruth Lundin. *Your Child's Speech and How To Improve it.* Cleveland: The Press of Western Reserve University, 1949, pp. 30. (paper) $0.75.

Children Absent from School. A Report and a Program. New York (136 East 57th Street): Citizens' Committee on Children of New York City, 1949, pp. 112. (paper) $1.00.

Lee J. Cronbach. *Essentials of Psychological Testing.* New York: Harper and Brothers, 1949, pp. 475. $4.50.

Maurice R. Davie. *Negroes in American Society.* New York: McGraw-Hill Book Co., 1949, pp. 542. $4.50.

Paul E. Fields. *Instructor's Manual To Accompany John Frederick Dashiell's "Fundamentals of General Psychology."* Boston: Houghton Mifflin Co., 1949, pp. 155. (paper) $3.40.

Clifford P. Froehlich and Helen E. Spivey. *Guidance Workers' Preparation.* A Directory of the Guidance Offerings of Colleges and Universities. Washington: U. S. Office of Education, 1949, pp. 45. (paper)

Kenneth G. Fuller. *An Experimental Study of Two Methods of Long Division.* New York: Bureau of Publications, Teachers College, Columbia University, 1949, pp. 76. $2.10.

I. H. Goldberger and William H. Bristow, Directors. *Physical Activities for Elementary Schools. A Manual for Physical Activities in the Health Education Program.* New York:

Board of Education of the City of New York, 1949, pp. 242. (paper)

FLORENCE L. GOODENOUGH. *Mental Testing: Its History, Principles, and Applications.* New York: Rinehart & Co., Inc., 1949, pp. 609. $5.00.

SUDHIN N. GHOSE. *And Gazelles Leaping.* New York: The Macmillan Co., 1949, pp. 238. $3.50.

ROBERT HOPPOCK. *Group Guidance: Principles, Techniques, and Evaluation.* New York: McGraw-Hill Book Co., 1949, pp. 393. $3.75.

RICHARD H. LAMPKIN. *Variability in Recognizing Scientific Inquiry: An Analysis of High-school Science Textbooks.* New York: Bureau of Publications, Teachers College, Columbia University, 1949, pp. 79. $2.10.

MERLE LAWRENCE. *Studies in Human Behavior.* Princeton, N. J.: Princeton University Press, 1949, pp. 184. (paper) $3.50.

THURBER H. MADISON, DOROTHY G. KELLEY, AND WILLIAM H. FOX. *Music in the Elementary Schools of Indiana.* Bloomington, Ind.: Indiana University, 1949, pp. 63. (paper) $0.75.

MARGARET MEAD. *Male and Female: A Study of the Sexes in a Changing World.* New York: William Morrow & Co., Inc., 1949, pp. 477. $5.00.

FREDERICK MOSTELLER, HERBERT HYMAN, PHILIP J. McCARTHY, ELI S. MARKS, AND DAVID B. TRUMAN. *The Pre-election Polls of 1948.* Report to the Committee on Analysis of Preelection Polls and Forecasts. New York: Social Science Research Council, 1949, pp. 396. (paper) $2.50.

LAWRENCE I. O'KELLY. *Introduction to Psychopathology.* New York: Prentice-Hall, Inc., 1949, pp. 736. $4.50.

MURIEL CATHERINE POTTER. *Perception of Symbol Orientation and Early Reading Success.* New York: Bureau of Publications, Teachers College, Columbia University, 1949, pp. 69. $2.10.

STAFF OF THE READING CLINICS OF THE UNIVERSITY OF CHICAGO. *Clinical Studies in Reading. I.* Chicago: The University of Chicago Press, 1949, pp. 173. (paper) $3.50.

WILLARD L. VALENTINE AND DELOS D. WICKENS. *Experimental Foundations of General Psychology.* New York: Rinehart & Co., Inc., 1949, pp. 472. $3.00.

J. E. WALLACE WALLIN. *Children with Mental and Physical Handicaps.* New York: Prentice-Hall, Inc., 1949, pp. 549. $5.00.

ELLIS WEITZMAN AND WALTER J. McNAMARA. *Constructing Classroom Examinations: A Guide for Teachers.* Chicago: Science Research Associates, 1949, pp. 153. $3.00.

ERNEST GLEN WEVER. *Theory of Hearing.* New York: John Wiley and Sons, Inc., 1949, pp. 484. $6.00.

ALFRED NORTH WHITEHEAD. *The Aims of Education.* New York: New American Library (245 5th Ave.), 1949, pp. 166. (paper) $0.35.

MARY HAYDEN BOWEN WOLLNER. *Children's Voluntary Reading as an Expression of Individuality.* New York: Bureau of Publications, Teachers College, Columbia University, 1949, pp. 117. $2.35.

ARTHUR WORMHOUDT. *The Demon Lover: A Psychoanalytical Approach to Literature.* New York: The Exposition Press, 1949, pp. 150. $3.50.

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 41 May, 1950 Number 5

AN EXPERIMENTAL PROJECTIVE TECHNIQUE FOR THE ANALYSIS OF RACIAL ATTITUDES*

GRANVILLE B. JOHNSON, JR.
Arizona State College at Flagstaff

The important rôle which attitudes play in the social development and integration of individuals and societies has precluded much significant research concerning it. The techniques of measurement and evaluation utilized in research on attitudes are as varied as the many problems intrinsic to the study.

A review of the recent literature on attitudes reveals that a large majority of investigations concerning attitudes uses the objective technique in which attitude scales are devised, applied to a specific group, and the results treated statistically. Each attitude is considered a specific characteristic of the subject and it acquires significance through comparison with other members of the selected population. There is most often no attempt to relate the revealed 'attitude' to the total individual. "The resulting conclusions from such a psychologically molecular analysis of attitudes have usually been that people have different attitudes on different subjects, that these attitudes are often internally inconsistent and irrational, that people acquire attitudes and stereotypes as a result of suggestion, propaganda, conditioning and fortuitous experiences which are imposed upon them by outside forces."[1]

Opposed to the molecular analysis of attitudes is the technique utilizing the molar approach. This method of investigation considers an attitude as a manifestation of a total reaction, a part which contributes to the whole and which is determined by the whole. In the study of racial antagonism, for example, it is not sufficient to classify the selected population into pro and con categories. The total personality must be considered as completely as clinical tools will permit in order to determine the bases for the prejudice.

---

* This article is the first of a series of three concerned with the origin and development of racial attitude. The second study, entitled "The Origin and Development of the Spanish Attitude Toward the Anglo and the Anglo Attitude Toward the Spanish," and the third, "The Relationship Existing Between Bilingualism and Racial Attitude," will appear in later issues of THE JOURNAL.

THE PROBLEM

The purpose of the present study was to devise an instrument which could be used in the investigation of racial attitude bases and development by analyzing personality as a whole, thus making it possible to investigate the dynamics of behavior.

MATERIALS AND SUBJECTS

Materials.—(1) The Projective Test of Racial Attitudes composed of six carefully selected pictures relevant to race situations and provocative of projected response was used for the analysis of racial attitudes. There was a hero figure with whom the subject identified himself and thus projected something of the dynamics of his personality into the depicted situations.

The situations conducive to projected response were identical except for the ethnic group variable. For Anglo subjects there were six cards of total Anglo content and six with an Anglo hero in a Spanish situation. For Spanish objects the situations were reversed with a Spanish hero in an Anglo situation in six cards with identical cards without the Anglo element.

(2) A tabulation sheet was used for recording the dynamics of responses.

Subjects.—(1) Anglo: thirty male subjects (four-, eight- and twelve-year levels, N = 90) from a Southwestern public school system.

(2) Spanish: thirty bilingual male subjects (four-, eight- and twelve-year levels, N = 90) from the same Southwestern town.

A further limitation was that members of both groups were required to have lived in this town for the year prior to testing.

THE PROJECTIVE TEST FOR THE STUDY OF RACIAL ATTITUDES

The present investigator through empirical analysis decided that the most successful method of investigating attitude bases and development was by use of the projective technique which makes it possible to analyze personality on a molar level. This technique distinguishes between opinion and attitude and makes it possible to investigate the dynamics of behavior; these are two factors not found adequately in the traditional objective techniques.

The Selected Pictures of the Projective Test of Racial Attitudes.—When the Projective Test of Racial Attitudes was originally conceived, eighteen pictures were experimentally tested. It was found that the six in the present series were adequate in that they covered generally most social situations an individual of the selected ages would ordinarily meet; they fell within the attention span of even the four-year-olds, were of intrinsic interest at each of the selected age levels, and contained life situations with portrayed behavior most easily empathized by the subjects.

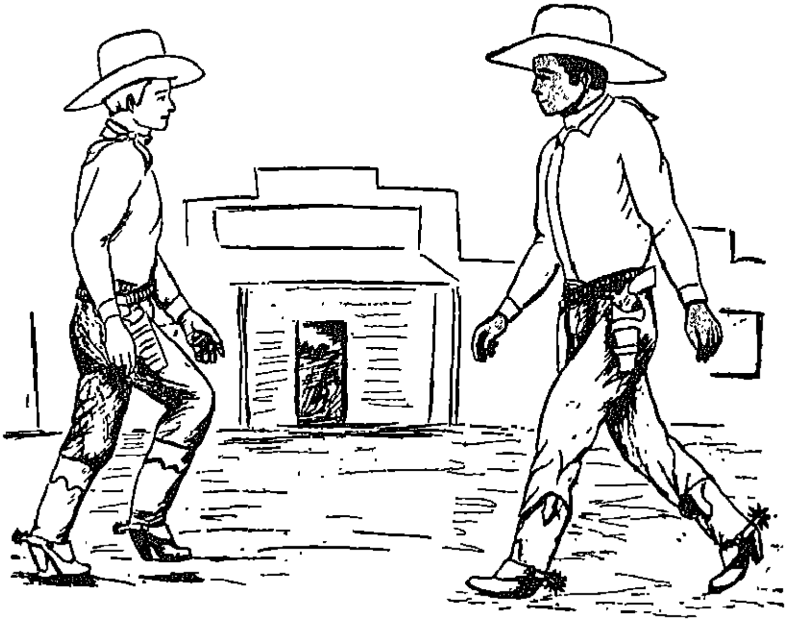


Fig. 1.

The cards are in color and consequently cannot be reproduced here. However, one is reproduced in black and white to show the character of the pictures. (See Fig 1.) They were presented in the following experimentally derived sequence:

For Spanish Subjects—

1) Spanish and Anglo cowboys approaching each other.
2) Spanish boy running down the street.
3) Spanish boy and Anglo boy holding a bat.
4) Two Spanish boys standing with broken toy between them.
5) Anglo girl crying with Spanish boy observing.
6) Spanish girl possibly inviting Spanish boy to share soda.
7) Two Spanish cowboys approaching each other.
8) Anglo boy running down the street.
9) Two Spanish boys holding bat.
10) Spanish boy standing with Anglo boy with broken toy between them.
11) Spanish girl crying with Spanish boy observing.
12) Anglo girl possibly inviting Spanish boy to share soda.

For Anglo Subjects—

1) Anglo and Spanish cowboys approaching each other.
2) Anglo boy running down the street.
3) Anglo boy and Spanish boy holding a bat.
4) Two Anglo boys with broken toy between them.
5) Spanish girl crying with Anglo boy observing.
6) Anglo girl possibly inviting Anglo boy to share soda.
7) Two Anglo cowboys approaching each other.
8) Spanish boy running down the street.
9) Two Anglo boys holding a bat.
10) Anglo boy standing with Spanish boy with broken toy between them.
11) Anglo girl crying with Anglo boy observing.
12) Spanish girl possibly inviting Anglo boy to share soda.

Administration.—

A) Physical conditions: The physical conditions of the test situation were necessarily controlled. Good lighting was essential because identification by the subject of the characters in the pictures was of utmost importance. It was important, too, to have sufficient table space with no inhibitory stimuli in evidence.

Rapport between the subject and the administrator was necessarily established before testing could proceed. Friendliness and assurance were important in the administrator.

B) Instructions: After rapport had been established, the following information was secured from the subject: (1) age, (2) name, (3) residence in the town (minimum of one year). At this point nothing was mentioned concerning Spanish and Anglo relations.

The directions for administration are as follows: "I am going to show you two pictures (present the Spanish and Anglo stereotypes for mental set) and I want you to tell me which of the two boys you like better—which would you rather be?" Then, "Which do you look like?" After the choice had been made, the following directions were given: "This is not a test. I am just going to show you some pictures and you are going to make up a story about each one—an exciting story—as exciting as you can. There are no right or wrong answers. You have all the time you want. All you are going to do is to make up an exciting story about the pictures I am going to show you."

After the administrator had answered any questions, the first picture was presented.

C) Procedure: 1) Time. (a) The reaction time was recorded. (b) The total time was recorded.

2) Recording. In the upper section of the answer sheet pertinent information was recorded regarding age, name, length of residence in the town with space for further information. The answer sheet was divided into two columns. The responses to the first six pictures were recorded in essence, and verbatim, if particularly significant, in the left column, and the second six in the right column. In this manner, the Spanish-Spanish (or Anglo-Anglo) pictures were directly across from their Spanish-Anglo counterparts. Roman numerals signified the pictures. The reaction time was recorded directly behind the numeral and the total time immediately after the story.

3) Questioning. When necessary inquiry was made concerning the incidents which lead up to the situation depicted, what was happening at the present time, and what would ultimately happen. Questioning was subtle and not leading. It was necessary to repeat that an exciting story was desirable.

After stories had been told for each of the twelve pictures, the subject was asked in direct questioning, how he got along with the members of the other ethnic group; in which were most of his friends; and what he thought of the other ethnic group in general. From these data, a correlation between opinion and attitude was made.

4) Length of session. The time required to administer the test generally ranged from thirty to sixty minutes depending on the intelligence, imagination, and stability of the subject.

Principles of Interpretation.—The stories were first read as a whole so as to determine roughly the general trends in the themes.

The method of scoring employed an objective type chart.[2]

- I. Effect of the environment on the organism.
  - A) Frustrating: rejection, withdrawal.
  - B) Neutral.
  - C) Helpful: acceptance, friendliness.
- II. Reaction of the organism to the environment.
  - A) Frustration: insufficiency.
    - 1. Aggression.
    - 2. Submission.
  - B) Self-sufficiency: emotional stability.
    - 1. Attainment.
    - 2. No conflict.
- III. Adequacy of the principal character.
  - A) Superordination of hero.
  - B) Equality.
  - C) Subordination of hero.
- IV. Ending (hero).
  - A) Defeat.
  - B) Indeterminate.
  - C) Victory, success.
- V. Ending (theme).
  - A) Unsatisfactory to society.
  - B) Indeterminate.
  - C) Satisfactory to society.

After the test was administered, it was scored with respect to the four categories, and the chart for scoring, Table I, was marked. Since each picture was identical with its prime number picture except for the race variable, any difference between the 'Spanish-Spanish' and 'Spanish-Anglo' columns were attributed to the single variable (order of administration was controlled). This was done by finding the standard error of the per cent for the total for each year and then finding the critical ratio. The size

TABLE I.—TABULATION SHEET

Bilingual Score__________

Name__________Age____Grade_____Ethnic Group________Residence__________

	Effect of Environ.		React. to Environ.		Adeq. of Pr. Char.		Ending (Hero)		Ending (Theme)	
	Sp-Sp or Sp-Ang		Sp-Sp or Sp-Ang		Sp-Sp or Sp-Ang		Sp-Sp or Sp-Ang		Sp-Sp or Sp-Ang	
	Ang-Ang F N H	F N H	Ang-Ang F S	Sp E Sb	Ang-Ang Sp E Sb	Sp E Sb	Ang-Ang D I V	D I V	Ang-Ang U I S	U I S
Picture 1-1′	A B C	A B C	A B	A B	A B C	A B C	A B C	A B C	A B C	A B C
2-2′	A B C	A B C	A B	A B	A B C	A B C	A B C	A B C	A B C	A B C
3-3′	A B C	A B C	A B	A B	A B C	A B C	A B C	A B C	A B C	A B C
4-4′	A B C	A B C	A B	A B	A B C	A B C	A B C	A B C	A B C	A B C
5-5′	A B C	A B C	A B	A B	A B C	A B C	A B C	A B C	A B C	A B C
6-6′	A B C	A B C	A B	A B	A B C	A B C	A B C	A B C	A B C	A B C
Total										

Reaction Time—Theme

1		7	
2		8	
3		9	
4		10	
5		11	
6		12	
Total			

Total Time—Theme

1		7	
2		8	
3		9	
4		10	
5		11	
6		12	

of the difference was assumed to be determined by and reflective of the strength of the stimulus to the particular subject. The strength of the stimulus varied from subject to subject since each experience was affect-provoking to the degree that it reflected the individual's total adjustment pattern.

The findings of the Projective Test of Racial Attitudes were then compared with the degree of bilingualism; that is, differences for each of the five categories were compared with the extent of bilingualism.

Scoring the Projective Test of Racial Attitudes.—It is important in the interpretation of protocols derived from projective techniques to ascertain whether or not the subject is identifying himself with one of the characters in his stories since he sometimes does not. In some cases, he manifests an attitude or theory regarding the outside world.

Five principles may be used in determining when the subject is identifying, and with which character in his story:

(a) Other conditions being constant, the subject is likely to identify himself with a character of the same sex, either one of his own age or a previous age.

(b) Other conditions being constant, the subject is likely to identify himself with the central character of the story, that is, the one around whom the story revolves, whose feelings are being expressed, whose behavior is being described, the one who is likely to be described first and figures in the ending.

(c) Other conditions being constant, the subject is likely to identify himself with a character who does not have behavior which is socially unacceptable 'in the subject's eyes.'

(d) The subject is more likely to be identifying with one of the characters when he gives evidence of emotional involvement in the story he tells.

(e) The subject is likely to identify himself with a character whose history is most similar to his own.[3]

In the Projective Test of Racial Attitudes, it was decided that the ethnic group to which the subject belonged determined the hero or principal character. In the Spanish-Spanish and Anglo-Anglo situations, the dominant individual was considered the hero as reflected in the category 'adequacy of the principal character.' (Adequacy of the character with whom the subject identified himself.) Because of the manner in which the scoring chart was set up, this technique was considered justifiable. Since the present investigation was concerned with prejudice, consistent identification of the subject with one or the other ethnic groups was significant in light of the five principles of identification outlined above.

Following is the frame of reference for scoring the Projective Test of Racial Attitudes. This scoring key was set up before any of the protocols were analyzed. It was necessary, however, to add to it when new problems arose.

KEY FOR SCORING THE PROJECTIVE TEST OF RACIAL ATTITUDES

The hero of the Spanish-Spanish or Anglo-Anglo pictures is the dominant individual. The hero of the Spanish-Anglo situations is the ethnic group to which the subject belongs. If the protocol is divided into a complete Anglo-Anglo and a Spanish-Anglo story, grade both categories. When extra scapegoat (minority group, not including Indians), grade from standpoint of principal character and from standpoint of attitude toward minority group.

Effect of Environment.—Primarily the environment leading up to the present situation.

Helpful: In the Anglo-Anglo or Spanish-Spanish situation, one betters the position of the other. In the Spanish-Anglo situation, the Spanish must help the Anglo or the Anglo must help the Spanish. In the Anglo-Anglo or Spanish-Spanish boy-girl situation, the girl must better the boy's position. If the two characters are in business together—helpful.

Frustrating: Does not tend to set up a problem for the other? Does the Spanish set up a problem for the Anglo in the Spanish-Anglo situation? Hindrance.

Neutral: Neither helpful nor frustrating.

Reaction to the Environment.—Is the effect of the environment frustration provoking or not? Frustration is non-integrative behavior, e.g., fight, flee, or some other overt expression of non-integrative emotion.

Adequacy of Principal Character.—Assume that the ethnic group of the subject is the hero.

Superior: Where the hero's will is imposed on the environment. If in Card II the subject puts the object in a frustrating situation that is superior. In games, the person in most desirable position from subject's standpoint is superior.

Equal: No conflict of will.

Subordinate: Will of other member of the environment is imposed on the hero. Is the hero superior, equal, or subordinate in the environment set up by the other member?

If the boy hits the girl for fun, that is frustration. If he is chastised for act, that is subordination and defeat if the girl is not punished too. Physical, mental, skill, or any other feature is criterion of adequacy.

Explicit behavior alone is graded except when the subject qualifies behavior. Buying something for girl is not superior.

Defeat in fight equals subordinate. Flees and gets away equals superior. One wins in fight one day and loses next equals equal. Always use one category if possible. If one wins in one thing and loses in another equals superior and subordinate. If one wins in one thing and loses in same thing equals equal.

Ending Hero.—Does the hero end victoriously or defeated with reference to the other member of the picture?

Non-aggression is indeterminant.

If ethnic group I (Hero) has fight with ethnic group II, and wins, that is Victory. Then, if they are both punished for fighting by some member of ethnic group I, that remains Victory. If punished by some member of ethnic group II, that is Defeat.

Ending Theme.—Does the story turn out socially integrated with reference to the mores of society (necessary here to project). Transgression yields retribution. Is the super-ego developed as per the mores of the social group?

Adherence to this frame of reference is essential in the analysis of the protocols.

STATISTICAL ANALYSIS OF THE PROJECTIVE TEST OF RACIAL ATTITUDES

Reliability.—Probably the most valid measure of reliability of scoring technique is to compute the degree of correlation between judgments of two or more trained analysts. In the present investigation, the protocols were scored independently by the investigator and another clinician trained in projective methodology.

TABLE II.—RELIABILITY OF SCORING THE PROJECTIVE TEST OF RACIAL ATTITUDES USING SIXTY SPANISH AND SIXTY ANGLO SUBJECTS DERIVED BY CORRELATING JUDGMENTS OF TWO ANALYSTS

	4 yrs.		8 yrs.		12 yrs.	
	r	σr	r	σr	r	σr
Anglo	.9859	.0031	.9724	.0050	.9691	.0045
Spanish	.9679	.0029	.9466	.0047	.9367	.0111

Table II is a presentation of reliability of scoring technique ascertained by correlating judgments on one hundred and twenty protocols of two trained analysts.

It is noted from Table II that the Spanish protocols yielded slightly less consistency in judgment than the Anglo; and the protocols of the older subjects elicited greater error in scoring than the younger. The former finding was possibly due to the subjectivity (projection) of the two analysts who were both members of the Anglo ethnic group. The latter was probably the result of increased complexity of stories elicited in the older age groups.

A more thorough analysis of scoring consistency was made to discover the reliability of each of the columns in the scoring chart.

The findings indicate no consistent variability. It is the opinion of the present investigator that the derived coefficients of reliability were, to a marked degree, attributable to the effectiveness of the frame of reference and would have approached perfect agreement if the scoring key had been adequately used.

EFFECT OF ETHNIC GROUP OF ADMINISTRATOR ON TEST RESULTS

After the reliability of the Projective Test of Racial Attitudes was established, the effect on test results of the ethnic group to which the administrator belonged was determined. It was realized that the results might be colored by the administrator, particularly if he were a stereotyped member of either of the two groups under consideration.

After two weeks of in-service training including actual practice paralleled by instruction in theoretical background, a Spanish stereotype was allowed to administer a representative sample of tests. A larger sample would have been desirable; but, because of the limited number of available subjects, and the fact that results could possibly be colored by this variable, only one-third of the subjects at the eight-year level in each ethnic group were considered.

An individual with the physiological characteristics of the stereotyped Anglo administered the test to ten subjects in each ethnic group at the eight-year level, and the Spanish administrator to ten others. The difference between the ten Anglo administered tests and the ten Spanish administered tests though manifesting inconsistencies (probably because of small sample) demonstrated no constant variation.

As a preliminary experimental procedure, the Anglo investigator administered ten tests which were readministered by the Spanish administrator three months later. The time interval was to remove any effects of practice. The method was necessarily discontinued. The test re-test correlations, though varying in theme, were not altered to a great extent in scored form. The correlations with the small sample were over .80.

With reference to the foregoing analysis, the tentative conclusion was drawn that the ethnic group to which the administrator belonged was not an uncontrolled variable which must be considered in test interpretation.

The Prejudice Score.—It was found that, in order to study the characteristics of the individual protocol, a single score indicating extent of prejudice had to be ascertained. This was done by finding the per cent difference on the tabulation sheet between same group and not-same group categories and adding algebraically. The result was a 'prejudice' score. Analysis of the score, too, brought out the relationships existing between the categories of the scoring chart.

Since the prejudice score was composite of the differences between the per cents of deviate scores of the same group and not-same group categories, and represents a general factor, it was necessary to measure the significance of the contribution made by each difference. By 'deviate scores' is meant those categories which are not neutral, equal, or indeterminate; these latter being non-reflective of positive or negative attitude.

Coefficients of correlation were between each category contributing to the prejudice score minus the weight of the variable category. This procedure was applied at each age level for each ethnic group.

It was noted that, though some categories contributed a negative value at some age levels, the weighted contribution was sufficiently small, (it was demonstrated by experimentation) that little bearing on prejudice score was in evidence.

The prejudice score proved invaluable in the analysis of other variables which demanded individual analysis.

Time.—

(1) Reaction time. The time elapsing between presentation of the stimulus and the subject's first reaction to that stimulus in word association, Rorschach, and other projective techniques, is considered quite important as an indication of emotional arousal and blocking.

One of the features of the Projective Test of Racial Attitudes was to measure the reaction time to the Anglo-Anglo (Spanish-Spanish) pictures and compare it with the reaction time to the Spanish-Anglo pictures. With reference to past use of reaction time, it might be expected that, reflective of increased affect and blocking, the reaction time to the Spanish-Anglo pictures would be significantly greater than reaction time to the like-group situations.

Analysis of Spanish subjects revealed no significant relationship between prejudice and reaction time. Tests of curvilinearity demonstrated no significant deviation from rectilinearity, though there appeared to be a propensity for the level of significance of curvilinearity (x^2Nyx) to increase with age. This would indicate some tendency for Spanish subjects of high and low reaction time to have less prejudice.

In analysis of Anglo reaction time correlated with the prejudice score, it was observed that, though none of the rectilinear correlations are significant, they are uniformly negative. This would indicate that there is a slight though consistent tendency to react more rapidly to emotion provoking stimuli. This is particularly manifest at the twelve-year level.

Analysis of curvilinearity reinforces the tentative conclusions derived with reference to the progressive tendency for those subjects with less prejudice to elicit greater or lesser reaction time. With reference to the relevant literature on the significance of reaction time, no indication of similar conclusions from previous research has been revealed.

(2) Total time. Total time appears particularly at the higher age levels, to be more reflective of emotionality than is reaction time. This conclusion was reached after computing the degree of correlation between prejudice score and total time.

It was noted that as age increases, there was greater relationship existing between the prejudice score and total time.

Analysis of curvilinear relationships revealed no significant deviation from rectilinearity.

(3) Reaction time and total time. In an attempt to reveal more of the characteristics of reaction time and total time, the relationship existing between the two was determined.

The rectilinear correlation at the four-year level alone is significant (r = .632). This would indicate that there is a definite rectilinear relationship between reaction time and total time only at this level.

Curvilinear analysis shows that there is a tendency at the four-year level for those subjects with high and low reaction time to have the least total time. This characteristic, however, becomes progressively less manifest.

Probably the most outstanding feature revealed by the curvilinear analysis is the significant tendency (Eta = .881) at the twelve-year Spanish level for subjects with high and low total time to have low reaction time. This appears to be a recently acquired characteristic of the twelve-year Spanish since there is little indication of this quality at the eight-year level. There is a possibility that this finding may be explained in the following manner—the subject is emotionally aroused and reacts in one of two ways: (1) By responding rapidly to the card, giving a perfunctory, non-revealing response and then going to the next card where he repeats the performance. This would, perhaps, be similar to the evasive nature of the 'cheap whole' to the Rorschach. This hypothesis is given added credence when it is found that the individual subjects are quite consistent in the length of time they require on the individual cards. (2) By immediately seeing the added 'exciting' possibilities in the Spanish-Anglo card and rapidly responding to it. This hypothesis was borne out by subjective evaluation.

The rectilinear correlations appear to be least significant at the four-year Anglo level, (r = −.04). This is the antithesis of the Spanish four-year level where the relationship between reaction time and total time was most significant of the age levels, (r = .63).

There is also indication, but at a less significant degree than that found in the Spanish subjects, for Anglo subjects with high and low total time to have low reaction time. The possible explanation for this phenomenon offered for the Spanish subjects may also be applied to the Anglo.

With reference to the findings on reaction time and total time, it may be concluded that reaction time has no significance as an index of prejudice. Total time, however, is more significant, particularly at the higher age levels.

IDENTIFICATION OF SUBJECTS WITH STEREOTYPES

One of the features of administration of the Projective Test of Racial Attitudes was to present the pictures of stereotyped Spanish and Anglo boys and to ask the subject which one he most resembled. This was done to ascertain the rôle of ego in the development of group consciousness and group identification. It was surmised, too, that this procedure might throw light on the hero identification in the stories elicited. As has been pointed out, previous research indicates that ability to identify stereotypes is positively related to the extent of prejudice.

For the purpose of analyzing this feature of the Projective Test of Racial Attitude, the identification of the Spanish and Anglo subjects with the stereotyped Spanish and Anglo drawings was correlated with the individual prejudice score. This was done to ascertain the relationship between identification and attitude.

It is noted that the Spanish and Anglo subjects at the four-year level manifest equal ability to identify themselves with the ethnic group to which they belonged. The Spanish, however, appeared to learn more rapidly than the Anglo, the ethnic group to which they belonged. Only one case in each of the older groups failed to identify himself, though both failures were elicited from subjects who were typically Spanish. An analysis of these two isolated protocols brought out no significant qualities such as prejudice swing which would explain the subject's inability to identify himself correctly.

It was noted that the Anglos, as a result of the age variable, learned to identify themselves less accurately than the Spanish. This is possibly because the physical appearance of the Anglos is less homogeneous than the Spanish. This factor was considered while compiling the data and it was revealed that those older

Anglos with a high prejudice score who could most conceivably identify themselves with the Spanish stereotype failed to do so.

When the prejudice score was related with the self-identification of the Spanish four-year level, a significant correlation was obtained ($r = -.504$). The four-year level alone could be analyzed because the error of identification at the older age levels was insignificant.

It appears that the higher the prejudice score, the less prone the Spanish four-year-old was to identify himself with his ethnic group. This feature might be explained on the basis (which is borne out by subjective analysis) that the Anglo have a higher economic standing than the Spanish, hence can have more material objects such as better clothes and more toys. This desire appears to be coupled with resentment, as the coefficient of correlation would indicate. Further understanding of this subject may be gained upon analysis of the Anglo four-year level.

It is noted at the Anglo level that there is a progressive tendency for valid self-identification to be related to prejudice until, at the twelve-year level the relationship is marked ($r_{bis} = .43$). Similar to the correlation manifest at the Spanish four-year level, there is a tendency for those four-year Anglos of most accentuated prejudice to identify themselves as members of the Spanish ethnic group, possibly for the same reason suggested for the Spanish behavior. This surmise is supported by a typical statement made by a four-year Anglo whose prejudice score was positive and approximated the mean for members of that age and ethnic group. Upon being asked which of the two stereotypes he most represented, he indicated the Spanish though he was a typical Anglo. He then volunteered the following statement: "I like those boys (indicating Anglo) because they (indicating Spanish) won't play with me. They (Spanish) play a lot better games than the others and are stronger, but they won't play with me because I'm not big and strong enough." The Anglo attribution of superior physical prowess to the Spanish is in evidence throughout the Anglo age levels. Where it is looked upon with admiration by the members of the younger age levels, however, it is considered with contempt by the older. Examples of this shift are as follows:

"I look like (indicating Anglo) and I'd rather be white; if I'd rather be mean and tough, though, I'd rather be (indicating Spanish). But I'd rather be white."

"Most Spanish have better physical ability than whites."

"The Mexicans don't treat you right—they fight all the time. They're tough because they mature earlier than whites, but they're not as smart as whites."

These samples were drawn from the twelve-year Anglo level and the responses were volunteered. An analysis of these specific cases did not demonstrate any marked swing of prejudice: that is, this sensed superiority of the Spanish group was not reflected in greater or lesser bias on the part of the Anglo.

THE STEREOTYPED PICTURE LIKED BETTER

In an attempt to contribute further to the understanding of the prejudice factor, the subjects of both ethnic groups were asked to indicate the stereotype picture they liked better—which one they would rather be like. This choice was then correlated with the prejudice score.

A tendency was found in the Spanish ethnic group for those subjects at the four-year level who were more prejudiced, to like and want to be like the members of the Anglo group. Though this tendency is not significant, it is indicative. There appears to be a reversal in this attitude which is quite manifest by the eight-year level—both the eight- and twelve-year correlations are statistically significant in a position manner. This latter finding would indicate that those older Spanish subjects who preferred the Anglo stereotype and would rather have the appearance of the Anglo possessed a lower than average prejudice factor.

The relationship existing between the choice of stereotype with the prejudice score for the Anglo age levels was made. A comparison of the reaction at the Spanish four-year level with the Anglo four-year level shows a marked difference between groups. Where the prejudice score manifested a negative correlation at the four-year Spanish level, which indicated that those subjects with the greatest prejudice against the Anglo liked and preferred to be like the Anglo, there was a reversal of this finding at the four-year Anglo level which approached statistical significance (critical ratio at .10 level of significance). This would indicate that the Anglos at each age level who preferred the Spanish stereotype were the least prejudiced.

Opinion.—As has been demonstrated in previous research, opinion may not be reflective of attitude. Therefore, it was one of the areas of the current investigation to analyze any relationship which existed between opinion and attitude.

An inconclusive tendency was manifest by the Spanish to become more moderate of their opinion of the Anglos with increase in age. A decrease in positive and negative reaction as maturation took place is observable. This trend toward equality, however, does not appear in the Anglo group. Here, dislike of the Spanish increased markedly from the four- to eight-year level with a concommitant decrease in columns headed 'Liked Spanish-Anglo Equally.'

The next step in the analysis of the relationship existing between opinion and attitude was to correlate expressed opinion with the prejudice score of the Projective Test of Racial Attitudes.

At the Spanish four-year level, there is a negative relationship between opinion and attitude, ($r = -.20$), but by the eight-year level little relationship exists, ($r = .096$). Evidently, the significant positive correlation at the oldest age level, ($r = .71$), is the product of recent development.

The same trend was manifested when the Anglo age levels were analyzed. The tendency, however, was not as accentuated as that found in the Spanish group.

Analysis of older groups would be important to ascertain the relationship between opinion and attitude at more sophisticated levels.

INTERNAL CONSISTENCY OF THE PROJECTIVE TEST OF RACIAL ATTITUDES

Each card in the Projective Test of Racial Attitudes was selected on the basis of its ability to cover a particular area of the subject's experience range. The environmental variability was quite wide, and varied in complexity.

The contribution which each picture makes to the end result might be valuable in determining its effectiveness, but a picture that has a low correlation with the group might be invaluable because it is loaded with an important but esoteric factor. In the Thematic Apperception Test, as has been previously indicated, it has been found that the responses to the pictures cannot be inter-correlated because the pictures are uneven in the type and amount of response they elicit and because the method includes isolating factors from context.

This rationalization is, no doubt, true, but the present investigator was interested in (1) noting the existence of a general factor and (2) ascertaining the consistency of behavior from situation to situation with reference to racial attitudes. The technique employed intercorrelates the response to each picture with the responses to the other five for each of the ten categories for each ethnic group at each age level. The aim, then, of correlating each picture with the others was to note the existence of a general factor and to determine the consistency of subject reaction.

When the study had been accomplished and the results analyzed it was concluded that each picture of the Projective Test of Racial Attitudes had a positive relationship with the other pictures of the test. The degree of relationship, though differing, displays no consistent variation resulting from the influence of age and ethnic group. Though there is a marked general factor, each picture makes a specific contribution to the total prejudice score.

SUMMARY

Probably the most successful method of investigating attitude bases and development is by utilization of the projective technique which makes it possible to analyze personality on a molar level.

Six pictures were experimentally selected and proven to be adequate. They constitute the Projective Test of Racial Attitudes.

Techniques of administration and interpretation were standardized and rigidly controlled to eliminate uncontrolled variables which would invalidate results.

Reliability of scoring the Projective Test of Racial Attitudes was found to be high at each age level for both ethnic groups.

The effect on test results of the ethnic group to which the administrator belonged was analyzed. Upon analysis of this problem, the tentative conclusion was drawn that the ethnic group to which the administrator belonged had little bearing on test results and was not an uncontrolled variable which must be considered in test interpretation.

It was found that, in order to study the characteristics of the individual protocol, a single score indicating extent of prejudice had to be ascertained. This was accomplished by finding the per cent difference of the tabulation sheet between same group and not-same group categories and adding algebraically. The result was a 'prejudice' score, which proved invaluable in the analysis of other variables which demanded individual analysis.

One of the features of the Projective Test of Racial Attitudes was to measure the reaction time to the Anglo-Anglo (Spanish-Spanish) pictures and compare it with the reaction time to the Spanish-Anglo pictures. No significant rectilinear or curvilinear relationship between racial attitude and reaction time was noted, but there was a tendency in both ethnic groups for those subjects with less prejudice to elicit greater or lesser reaction time. No indication of similar conclusions from previous research has been revealed.

Total time appeared, particularly at the higher age levels, to be more reflective of emotionality than was reaction time. This conclusion was reached after correlating prejudice score with total time.

The relationship between reaction time and total time was determined and it was found that there was a tendency for those subjects with high and low total time to have low reaction time.

One of the features of administration of the Projective Test of Racial Attitudes was to present the pictures of stereotyped Spanish and Anglo boys and to ask the subject which one he most resembled. It was noted that the Anglos, as a result of the age variable, learned to identify themselves less accurately than the Spanish. This is possibly because the Anglo physical appearance is less homogeneous than the Spanish. It was revealed, however, that those older Anglos with a high prejudice score who could most conceivably identify themselves with the Spanish stereotype failed to do so. There was indication that the higher the prejudice score, the less prone the four-year-olds of both ethnic groups were to identify themselves with their own ethnic group. A progressive tendency was manifest in both ethnic groups for valid self-identification to be related to prejudice score until, at the twelve-year level the relationship was marked.

A tendency was found for those Spanish subjects at the four-year level who were most prejudiced, to like and want to be like the members of the Anglo group. Though this tendency was not significant, it was indicative. There appeared to be a reversal in this attitude which was quite manifest by the eight-year level—both the eight- and twelve-year correlations were statistically significant in a positive manner. This latter finding would indicate that those older Spanish subjects who preferred the Anglo stereotype and would rather have the appearance of the Anglo possessed a lower than average prejudice factor.

A comparison of the reaction at the Spanish four-year level with the Anglo four-year level showed a marked difference. Where the prejudice score manifested a negative correlation at the four-year Spanish level, which indicated that those subjects with the greatest prejudice against the Anglo liked and preferred to be like the Anglo, there was a reversal of this finding at the four-year Anglo level which approached statistical significance (critical ratio at .10 level of significance). This would indicate that the Anglos, at each age level, who preferred the Spanish stereotype were least prejudiced.

Since opinion may not be reflective of attitude, one of the areas of the current investigation was to analyze any relationship existing between opinion and attitude. An inconclusive tendency was manifested by the Spanish to become more moderate of their opinion of the Anglos with increase in age. A decrease in positive and negative reaction as maturation took place was observable. This trend toward equality, however, did not appear in the Anglo group. Here, dislike of the Spanish increased markedly from the four-year to the eight-year level with a concomitant decrease in columns headed 'Liked Spanish Better' and 'Liked Spanish-Anglo Equally.'

The next step in the analysis of the relationship existing between opinion and attitude was to correlate expressed opinion with the prejudice score of the Projective Test of Racial Attitudes. It was found that, at the four-year level in the Spanish group, there was a marked negative relationship between the two factors under consideration, but by the eight-year level little relationship existed. Evidently, the significant positive correlation at the oldest age level was the product of recent development.

The same trend was manifested when the Anglo age levels were analyzed. The tendency, however, was not as accentuated as that found in the Spanish group.

Each card in the Projective Test of Racial Attitudes was selected on the basis of its ability to cover a particular area of the subject's experience range. The environmental variability was quite wide, and varied in complexity.

The present investigation was concerned with, (1) noting the existence of a general factor, and (2) ascertaining the consistency of behavior from situation to situation with reference to racial attitudes. The technique employed intercorrelates the response to the other five for each of the ten categories for each ethnic group at each age level. After this procedure was followed, it was concluded, that each picture of the Projective Test of Racial Attitudes had a positive relationship with the other pictures of the test. The degree of relationship, though differing, displayed no consistent variation resulting from the influence of age and ethnic group. Though there was a marked general factor, each picture made a specific contribution to the total prejudice score.

REFERENCES

1) David Krech. "Attitudes and Learning: A Methodological Note," *Psychological Review*, Vol. 53, No. 5, September, 1946, p. 291.

2) The idea of scoring the thematic responses in this manner was procured from Ruth Clark (Genetic Psychology Monographs, 1944, Vol. 30, p. 47). H. A. Murray, in a recent letter to the present investigator, made some specific suggestions, and the present investigator, after advice from L. I. O'Kelly, and through experimentation, developed the scoring system in its present form.

3) Julian R. Rotter. "Thematic Apperception Tests: Suggestions for Administration and Interpretation," *Journal of Personality*, Vol. 15, September, 1946, p. 83.

ERROR SCORES AS A MEASURE OF CAREFULNESS

BENJAMIN FRUCHTER*
University of Texas

I. INTRODUCTION

There are a number of occupations that require extremely careful and meticulous work. An example is the navigator of the modern long-range bomber. He may be called upon to plot a course to hit a tiny dot of land in the broad Pacific. Miscalculation may result in the loss of expensive equipment and the lives of himself and the other crew members.

In order to test the hypothesis that there is a trait of carefulness and that a factor of carefulness can be isolated and identified, four tests were constructed. These are paper-and-pencil tests similar to some of the tasks of the navigator such as careful measuring and plotting on maps and charts and reading values from complicated scales.

These tests were administered to three hundred fifty-four unclassified aviation students in January, 1945. The directions did not stress either speed or accuracy to the exclusion of the other. Two scores were obtained from the same administration of each test—the number of correct responses and the number of errors. It was observed that the error scores had considerable range and variability. It was also observed that when the correct and error scores of a test were correlated the coefficients were relatively low, indicating considerable independence of the two scores. This suggested the possibility that the error scores might be functionally different from the correct-response scores and should be treated separately in an analysis. In order to examine the hypothesis that these tests measured a common trait of carefulness it was determined to submit them to factor analysis along with a number of reference tests.

II. THE TESTS†

1–2. Directional Plotting.—This is a test of the speed and accuracy with which an examinee can locate points and estimate

---

* The writer wishes to express his appreciation to Roger W. Birkman and William L. Grafton for assistance with the factor analyses reported in this study.

† For a more complete description of the tests see J. P. Guilford (Ed.),

directions on a chart. The coördinates of two points on a 30 × 30 chart are given. The task is to (1) locate the positions of the two points without making any marks on the chart, (2) determine the direction of the second position from the first in terms of a twelve-point compass rose shown at the side of the chart. The rights score is variable 1 and the wrongs score is variable 2.

3–4. Complex Scale Reading.—This test consists of a chart containing seven vertical scales. For each item the examinee is given values on two of the seven scales. By placing a straight edge through these points, the examinee is to read the corresponding value on a given one of the remaining five scales. The rights score is variable 3 and the wrongs score is variable 4.

5–6. Plotting.—This test consists of a 25 × 25 chart and a compass rose. The examinee is given a starting point in terms of coördinates and told to move so many squares west, so many squares southeast, so many squares east, etc., and then to read the final position in terms of coördinates. The rights score is variable 5 and the wrongs score is variable 6.

7–8. Plotting Accuracy.—This test is a variation of the Plotting test. The same kind of chart is used, but three orientation compasses are shown as compared with one in the plotting test. The points on these are marked with letters rather than directions. The examinee is instructed which compass rose to use in each item. The rights score is variable 7 and the wrongs score is variable 8.

In order to make up a matrix for analysis eleven tests, selected from the classification test battery routinely administered to all aviation students in the AAF, were intercorrelated with the eight scores (four rights scores and four wrongs scores) obtained from the carefulness tests. The list includes several apparatus classification tests, because they showed high correlations with the carefulness tests. A brief description of each of these reference tests follows:

9. Rotary Pursuit.—This is a modification of the Koerth Pursuit Test. The examinee tries with his right hand to keep a prod in contact with a metallic spot on a phonograph-type disc which is rotating at the rate of one revolution per second. While

---

Printed Classification Tests, Washington: United States Government Printing Office, 1947, pp. 680–686.

doing this, there is a second simultaneous task for the left hand which requires the examinee to keep one of two keys closed in correspondence with one of two lights.

10. Two-hand Coördination.—This test consists of a rotating black disc on which is a small brass 'target' disc. The target disc follows an irregular path. The task of the examinee is to keep a microswitch in contact with the target disc as long as possible. The movement of the microswitch is controlled by turning two handles similar to those on a lathe. The handles turn in the opposite direction from those of a conventional lathe, so that those having had previous lathe experience will not have an undue advantage.

11. Complex Coördination.—This is a serial, choice-reaction-time test in which each stimulus is one of thirteen spatial patterns of three lights each. In correspondence with each stimulus pattern, the correct response is a unique adjustment of imitation stick-and-rudder controls.

12. Rudder Control.—In this test the examinee sits in a mock cockpit of an airplane. His own weight throws the seat off balance unless he applies correction by means of a rudder-control mechanism. The score is the total time he keeps the cockpit pointed directly at a target light straight ahead.

13. Discrimination Reaction Time.—This test was designed as a test of speed of decision and reaction. There are four stimulus patterns, each consisting of a pair of lights, one red and one green. Corresponding to each stimulus pattern is a microswitch, the four switches being arranged in a diamond-shaped pattern. The position of each switch, upper, lower, right, and left, is associated with a corresponding direction of the red light with respect to the green light. The task of the examinee is to push the switch associated with a given lighted pattern. The score is the time it takes the examinee to push the correct switch.

14. Finger Dexterity.—This test consists of forty-eight square pegs in square holes. Each peg can be grasped by means of a thick circular button at its top. The examinee lifts each peg from its hole, turns it 180° clockwise and resets it in its hole. The score is the total number of pegs turned in the time allowed.

15. Speed of Identification.—This is a paper-and-pencil test in which the silhouette of a plane must be matched with the

identical silhouette in a rotated position. The stimulus silhouettes are presented in groups of four and the silhouettes with which they are to be matched in groups of five, one being a mislead. The test is printed directly on an IBM answer sheet.

16. Numerical Operations I.—This is a simple arithmetic test printed directly on an IBM answer sheet. It consists of simple problems of addition and multiplication with answers provided. The examinee's task is to mark the given answer as either correct or incorrect.

17. Mechanical Principles.—This test is similar to the well-known Bennett and Fry Mechanical Comprehension Test. The items are pictorial representations of simple mechanical principles. The answer choices are stated verbally.

18. Reading Comprehension.—This test is composed of eight paragraphs concerning which thirty-six questions are asked. The items were designed to test ability to make valid inferences from the reading material as well as ability to answer more direct questions about content.

19. Arithmetic Reasoning.—This test consists of thirty arithmetic-reasoning problems which require a minimum of formal mathematical training to solve.

III. THE ANALYSIS

The rights scores and wrongs scores were obtained from the same administration and were separated for analysis. Failure to do so would result in experimentally dependent variables due to the fact that both scores had been taken from the same set of responses.*

Two matrices were prepared for analysis. One contains the four rights scores of the carefulness tests plus the formula scores (R − .25W) of the eleven reference tests. The other contains the four wrongs scores plus the eleven reference tests. These reference tests had been analyzed several times previously and their factorial content for this population is relatively well established. They were included to facilitate cross-validation of factors in the two carefulness batteries as well as with factors isolated in previous studies with these reference tests.*

---

* For an analysis of these tests in a single matrix see J. P. Guilford (Ed.), *op. cit.*, pp. 686-694. In that solution four doublet factors, one for each carefulness test, were obtained. They are spurious and resulted from the fact that both rights, and wrongs scores for any given test were derived from the same sample.

The wrongs scores were reflected for correlational purposes. Intercorrelations of the carefulness tests and correlations of the carefulness tests with the reference tests are based on a sample of three hundred fifty-four unclassified aviation students. The intercorrelations of the reference tests were not available for this particular sample but are based on other comparable samples of unclassified aviation students with a total N of 1,920. The intercorrelations of all the variables involved in the two analyses are shown in Table 1.

Each matrix of intercorrelations was factored by the centroid method. Six factors each were extracted from the rights and wrongs matrices. These factors were rotated graphically to "psychologically meaningful positions" in accordance with the principles of simple structure and positive manifold. Inspection of the factor plots indicated little evidence of obliqueness and an orthogonal reference frame was maintained. Table 2 gives the rotated and centroid loadings for the rights battery and Table 3 the corresponding values for the wrongs battery.

IV. INTERPRETATION OF FACTORS

Rotated Factor I is interpreted in both batteries as a numerical-facility factor.

Test	Loading	
	Rights	Wrongs
Numerical Operations I	.78	.61
Complex Scale Reading	.57	.09
Plotting	.61	.05
Plotting Accuracy	.61	.18
Arithmetic Reasoning	.44	.51
Directional Plotting	.39	−.07

All the rights scores of the carefulness tests have numerical content. This is to be expected from the nature of the tasks.

---

* Ibid., chapter 28.

TABLE 1.—CORRELATION MATRIX*

Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Directional Plotting (R)																			
2. Directional Plotting (W)	48																		
3. Complex Scale Reading (R)	57	05																	
4. Complex Scale Reading (W)	10	28	43																
5. Plotting (R)	46	12	55	24															
6. Plotting (W)	07	20	13	39	42														
7. Plotting Accuracy (R)	46	09	51	20	71	19													
8. Plotting Accuracy (W)	04	14	14	34	26	41	46												
9. Rotary Pursuit	36	−10	37	−05	49	−12	48	−06											
10. Two-Hand Coordination	44	−04	44	01	51	−03	44	−04	33										
11. Complex Coordination	42	19	30	11	35	08	32	09	38	48									
12. Rudder Control	20	20	04	07	07	−01	05	−06	32	30	32								
13. Discrimination Reaction Time	46	14	43	12	32	08	35	13	18	23	36	07							
14. Finger Dexterity	16	03	17	00	24	05	22	05	33	27	35	11	26						
15. Speed of Identification	27	10	28	03	34	05	29	08	12	14	21	05	21	16					
16. Numerical Operations I	38	04	49	07	48	05	52	14	05	−04	05	−03	16	14	11				
17. Mechanical Principles	34	36	18	23	11	10	12	14	21	35	29	27	20	04	13	−16			
18. Reading Comprehension	37	14	34	09	19	08	23	11	03	05	16	−01	21	04	11	15	31		
19. Arithmetic Reasoning	45	17	46	16	32	08	34	16	−02	06	15	−01	23	01	04	35	27	50	

* Decimal points omitted. Wrongs scores are reflected.

TABLE 2.—RIGHTS SCORE BATTERY
(Decimal points omitted throughout)

Test	Rotated Loadings						Centroid Loadings						
	I	II	III	IV	V	VI	I	II	III	IV	V	VI	h^2
1. Directional Plotting (R)	39	28	44	38	09	20	75	—10	19	09	—12	06	63
3. Complex Scale Reading (R)	57	24	28	16	11	42	73	—27	—09	—06	—16	20	68
5. Plotting (R)	61	30	06	07	54	19	75	—14	—41	—16	06	—16	80
7. Plotting Accuracy (R)	61	24	10	08	50	16	73	—20	—34	—14	06	—17	74
9. Rotary Pursuit	13	60	02	07	17	29	52	33	—25	—08	—16	18	51
10. Two-Hand Coordination	05	50	18	10	49	05	58	37	—11	—17	13	—12	55
11. Complex Coordination	00	43	18	44	31	03	58	36	06	18	11	—07	52
12. Rudder Control	—08	52	12	17	—04	—12	26	44	10	08	—24	—11	35
13. Discrimination Reaction Time	14	13	22	44	17	27	52	—05	10	22	16	16	39
14. Finger Dexterity	11	37	—09	29	08	11	36	21	—18	22	10	05	27
15. Speed of Identification	11	00	03	22	26	25	34	—06	—09	13	24	12	22
16. Numerical Operations I	78	—04	—04	19	—03	—03	40	—54	—23	19	—28	—22	62
17. Mechanical Principles	—17	23	56	09	16	08	39	24	42	—22	09	07	44
18. Reading Comprehension	16	—09	58	11	06	17	40	—26	41	—15	07	10	43
19. Arithmetic Reasoning	44	—09	62	06	02	—02	47	—41	39	—18	—10	—14	61

TABLE 3.—WRONGS SCORE BATTERY
(Decimal points omitted throughout)

Test	Rotated Loadings						Centroid Loadings						
	I	II	III	IV	V	VI	I	II	III	IV	V	VI	h^2
2. Directional Plotting (W)	−07	−10	14	16	41	36	37	−18	−25	13	17	−29	36
4. Complex Scale Reading (W)	09	−01	05	00	14	60	38	−36	−32	−14	07	01	40
6. Plotting (W)	05	−05	−02	07	−07	64	31	−38	−31	−30	−09	−03	43
8. Plotting Accuracy (W)	18	−04	06	10	−15	57	34	−39	−15	−30	−15	04	41
9. Rotary Pursuit	12	56	−04	23	13	−14	33	50	13	−10	14	15	43
10. Two-hand Coordination	−09	65	28	13	−02	00	43	49	−08	06	−21	24	54
11. Complex Coordination	06	57	19	34	13	16	63	37	−03	−08	−04	07	55
12. Rudder Control	−04	42	00	−03	42	06	32	30	−22	12	31	13	37
13. *Discrimination Reaction Time*	18	*19*	*21*	41	00	13	48	08	19	−10	−12	−10	31
14. Finger Dexterity	18	32	−09	44	01	02	36	28	20	−30	05	−06	34
15. *Speed of Identification*	*04*	*10*	*07*	*34*	*01*	*07*	*28*	*10*	*10*	*−11*	*−07*	*−17*	*15*
16. Numerical Operations I	61	−01	−01	07	00	02	22	−24	43	−16	18	20	39
17. Mechanical Principles	−18	22	53	11	37	24	55	08	*−27*	42	*−09*	*−10*	58
18. Reading Comprehension	26	−11	57	18	08	05	41	−23	28	33	−19	−07	45
19. Arithmetic Reasoning	51	−06	56	−02	12	11	46	−37	33	34	−05	19	61

The wrongs scores do not have significant loadings on the numerical factor.

Rotated Factor II is interpreted in both analyses as a psychomotor-coördination factor.

Test	Loading	
	Rights	Wrongs
Rotary Pursuit	.60	.56
Rudder Control	.52	.42
Two-hand Coördination	.50	.65
Complex Coördination	.43	.57

Four of the six apparatus tests have significant loadings on this factor. A similar factor has been identified several times previously in analyses of the aircrew classification battery.* Several of the carefulness tests rights scores have loadings approaching significance on this factor, reflecting the high correlations of these tests with some of the apparatus tests.

Rotated Factor III is interpreted as a general-intellectual factor, possibly a combination of the general-reasoning and verbal factors usually obtained separately.

Test	Loading	
	Rights	Wrongs
Arithmetic Reasoning	.62	.56
Reading Comprehension	.58	.57
Mechanical Principles	.56	.53
Directional Plotting	.44	.14

Only one of the rights scores of the carefulness tests has an appreciable loading on this factor.

Rotated Factor IV is identified as a psychomotor-precision factor in both analyses.

---

* Ibid., p. 816.

Test	Loading	
	Rights	Wrongs
Discrimination Reaction Time	.44	.41
Complex Coördination	.44	.34
Directional Plotting	.38	.16
Finger Dexterity	.29	.44

A similar factor has been identified in previous analyses of these apparatus tests.* The Directional Plotting, rights score, *apparently involves some finger dexterity.*

Rotated Factor V, rights battery, is interpreted as a spatial-reference factor involving judgments of change of distance and direction. Rotated Factor V, wrongs battery, is identified as a visualization factor.

Test	Loading	
	Rights	Wrongs
Plotting	.54	−.07
Plotting Accuracy	.50	−.15
Two-hand Coördination	.49	−.02
Complex Coördination	.31	.13
Rudder Control	−.04	.42
Directional Plotting	.09	.41
Mechanical Principles	.16	.37

There is, so far as the writer knows, no precedent for the factor interpreted as spatial-reference and the identification is accordingly tentative. The loading of Directional Plotting, wrongs score, on the visualization factor is of interest. Presumably the ability to visualize the compass rose on the chart helps to keep down the error score on that test.

Rotated Factor VI, wrongs battery is interpreted as a carefulness factor. The corresponding factor in the rights battery is not interpreted since only one test has a large loading on it.

---

* Ibid., p. 317.

Test	Loading	
	Rights	Wrongs
Plotting	.19	.64
Complex Scale Reading	.42	.60
Plotting Accuracy	.16	.57
Directional Plotting	.20	.36

The wrongs scores of the four carefulness tests have significant loadings on the factor from the wrongs battery. Since a trait of carefulness was postulated in the construction of these tests the factor is so identified. The sixth factor from the rights battery has but a single carefulness test loaded on it. This, together with the fact that the second and third highest loadings on it are psychomotor tests (see Table 2), makes it unlikely that it is a carefulness factor too. The extent to which the carefulness factor is common to error scores in other tests remains to be determined.

V. DISCUSSION

The major question which this study has attempted to answer is whether the error score of a speeded test contains sufficiently unique variance to justify its treatment as a separate variable. The error scores of the four experimental tests included in this analysis did define a new factor. Had the error scores not been analyzed separately, the general conclusion would have been that no new factor resembling carefulness could be found in these carefulness tests.

Thurstone* has isolated a similar factor in his analysis of perceptual tests. The tests were scored either per cent wrong or per cent omitted. He says of this: "The factor involved may be concerned with accuracy or caution. More data should be available on a variety of test material for this factor before it can be identified. This finding does suggest, however, that the relative frequency of errors may represent a unique trait."

---

* L. L. Thurstone, "The Perceptual Factor," *Psychometrika,* 3, 1938, pp. 1–17.

The previous analysis of these data, referred to above, in which both the rights and wrongs were included in the same analysis, confirms the results that the error scores do yield a unique factor even when the rights scores from the same tests are included in the matrix.* Rights scores of the Plotting and Plotting Accuracy tests did show slight, but probably not significant loadings (.22 and .19, respectively) with the carefulness factor.

It may be assumed safely that, if correlations of rights and wrongs scores are not high, a fuller picture of the true functions measured by a test can be obtained by analyzing the scores separately than by analyzing formula scores. The results also imply that many an error has possibly been committed by combining rights and wrongs scores in the same formula. Unless the two are factorially similar the result may be very different than had been intended by the test-maker. In this connection, it is in order to suggest that if rights and wrongs scores from a test are considered in factor analysis or in composite predictions, they should be derived from separate forms in order to avoid spurious correlations.

Another result of interest is the high correlation of the carefulness tests with some of the apparatus tests. The carefulness tests had higher correlations with the apparatus tests than with the other paper-and-pencil tests with which they were correlated. This fact gives considerable encouragement to proponents of the belief that many factors appearing in apparatus tasks can be duplicated in printed tests.

SUMMARY

In order to test the hypothesis that a trait of carefulness could be identified, four tests similar to some of the tasks performed by the aerial navigator were constructed. These are paper-and-pencil tests of plotting on maps and charts and reading values from complicated scales. The tests were administered to three hundred fifty-four unclassified aviation students in the AAF and separate correct-response and error scores were obtained for each test.

It was observed that the error scores had considerable range and variability and low correlations with the correct-response scores. It was decided to intercorrelate and analyze the scores to determine whether a carefulness factor could be identified.

---

* Guilford, *op. cit.*, p. 691.

Eleven paper-and-pencil and apparatus tests which had been previously taken by the aviation students in the course of their classification were added as reference variables. Two matrices were prepared for analysis, one consisting of the correct-response scores of the carefulness tests plus the eleven reference variables and the other of the error scores plus the reference variables.

The two matrices were factored and the factors rotated to simple structure. Numerical (I), psycho-motor coördination (II), general-intellectual (III), and psycho-motor precision (IV), factors were isolated in both analyses. The rights battery yielded a spatial-reference factor for the fifth factor (R-V). The fifth factor on the wrongs battery (W-V) appears to be a visualization factor. The sixth factor on the rights battery had an appreciable loading on only one test and was not interpreted. The sixth factor on the wrongs battery (W-VI) is the carefulness factor whose isolation was sought.

The major conclusions are:

1) The error scores of the four experimental tests define a new factor.

2) Had the error scores not been analyzed separately no new factor resembling carefulness could have been found in these tests.

3) If correlations of rights and wrongs scores are not high, a fuller picture of the functions measured by a test can be obtained by analyzing the scores separately than by analyzing scores arrived at by scoring formulas such as R − .25W etc.

4) If rights and wrongs scores from a test are considered in composite predictions or factor analyses, they should be derived from separate forms in order to avoid spurious correlations.

CHOICE OF MAJOR SUBJECT AS RELATED TO AMERICAN COUNCIL EXAMINATION SCORE AND COLLEGE GRADES

AUDREY M. SHUEY

Assisted by Barbara Edgar
Randolph-Macon Woman's College

INTRODUCTION

While a number of investigators have reported a relation between intelligence and occupation, very few have been concerned with intelligence as related to the field of specialization at the college level. One psychologist, in attempting to determine whether or not intelligence increases from the freshman to senior year, presented some findings pertinent to this problem. After testing a Lawrence College freshman class on the 1934 edition of the American Council on Education Psychological Examination, Flory[1] grouped the students according to their choice of major subject and subsequently combined those in related fields. He found the Fine Arts group to rank first on the A. C. E., followed in order by the Language Arts group (English, French, German, Spanish, Latin, and Greek), the Science group (Biology, Chemistry, Geology, Mathematics, and Physics), and the Social Science group (Economics, Government, History, Psychology, Philosophy, Religion, and Sociology). Hartson[2] observed that Oberlin College freshmen who subsequently majored in the Mathematics—Science division scored higher on the average in numerical computations when tested by the Ohio State University Psychological Examination, while those freshmen who were to major in the Language-Literature groups showed a greater initial ability in handling verbal relationships.

The original purpose of the research to be reported in this paper was to determine whether or not women in a Liberal Arts College who major in one field differ in general intelligence from those who select another area for concentration. Finding that some departments did differ significantly from others in the general intelligence of their majors, we thereupon made a survey of grades by departments, since it seemed possible that students of somewhat lesser ability would have some difficulty in making their grades during their first two years and would select for further study those divisions of instruction where the grading was somewhat lenient. If this be the case one would not expect to find the various departments of different colleges ranking in the same way with respect to the general intelligence of their majors.

PROCEDURE

A score on the A. C. E. was used as the criterion of general intelligence in this study. For a number of years, the particular college studied has required all new students to take this test at a specific time during their first month of college residence. From the files of the Psychology Department were selected the names and test scores of all freshmen tested in the nine years between 1939 and 1947. The students tested in 1939 had been given the 1939 edition; those tested in 1940, the 1940 edition; . . . and those in 1947, the 1947 edition. Over the period of years they were all tested in the same large lecture rooms, at the same time of day, and by members of the Psychology Department. As far as the writers know, no new student who was on the campus during the testing period has ever been absent from the test.

Of the 2012 freshmen tested during the nine-year period, 1279 were in the college during their second year, selected a major field of concentration during April of this year, and indicated that they would return for the junior year. Of this number, a total of twenty-seven students selected German, Greek, Philosophy, Physics, or Religion for specialization, and were excluded from our survey because of the small numbers. The total gross scores of the remaining 1252 freshmen were all transmuted into the equivalent scores of the 1943 edition and categorized according to choice of one of thirteen major fields.

In determining average grades earned by students in the thirteen divisions of instruction under consideration, we included all students whose surnames began with A to L, inclusive, and who were in the college during their freshman and sophomore years between 1939 and 1947. For all divisions of instruction except Music and Art we have included only grades for six-hour courses, since the great majority of courses—other than in these two departments—in the first two years are year-courses allowing six semester hours of credit.* A preliminary study of 1340 consecutive grades of students entering college between 1939 and 1942 indicated that only thirty grades had to be omitted by this procedure. These thirty totalled sixty-three hours of credit as contrasted to the 7860 hours included in the 1310 grades. By our method of selection therefore we have probably included all but one per cent or less of the credits earned in the eleven divisions of instruction.† In Art and Music, all grades, regardless of hours' credit, were included with the exception of Choral Music which we excluded from our tabulations.

Ordinarily a student was given one final grade for the year's work in a course (six hours); however, if the two semesters' grades were recorded separately, the second grade, if not the same as the first, was arbitrarily taken as the grade. Furthermore, if a student dropped a given year course at the end of the first semester or at any other time with a record of D or F, the D or F was included just as if she had remained in the course throughout the year.

RESULTS

In Table I are included the number of majors in each of the thirteen divisions of instruction from 1939 to 1947, the mean gross scores and SD's of these students tested on the A. C. E. at the beginning of the freshman year, and the percentile ranks of these means as derived from the national norms. The A. C. E. means range from 133.50 for those students who later selected French as a major, to 119.07 for those who subsequently majored in Biology, the respective percentile ranks of the groups ranging from 86 to 68. The Music majors were most variable in their A. C. E. scores, as indicated by a standard deviation of 22.83, followed by English, History, and French; the SD's of the other nine groups were very similar, all of them being between 17 and 18.

---

* In the Art Department the courses are primarily four-hour courses. In Music likewise the great majority of classes open to freshmen and sophomores are four-hour courses.

† The omitted grades were from the following divisions of instruction: English, French, History, Latin, Mathematics, Political Science, and Spanish.

No significant differences were found among the freshman scores of groups who subsequently elected to major in French, Mathematics, Chemistry, Spanish, Art, Psychology, and Latin, at the one, two or five per cent level of confidence. Likewise none of the differences between the means of the groups who in their second year chose to major in History, English, Economics and Sociology, Political Science, and Biology proved to be significant at any of the three levels of confidence. However,

TABLE 1.—MEAN GROSS SCORES, STANDARD DEVIATIONS, AND PERCENTILE RANKS OF FRESHMEN TESTED BY THE A. C. E. FROM 1939 TO 1947 AND GROUPED ACCORDING TO SUBSEQUENT CHOICE OF MAJOR[1]

Major	N	Mean	SD	Percentile Rank[2]
Art	31	127.81	17.16	80
Biology	101	119.07	17.32	68
Chemistry	119	129.98	18.04	82
Economics & Sociology	151	120.93	17.74	71
English[3]	299	121.65	19.30	72
French	54	133.50	19.07	86
History	100	121.68	19.20	72
Latin	58	126.33	17.52	77
Mathematics	104	131.02	17.86	83
Music	33	126.00	22.83	77
Political Science	57	119.37	17.86	69
Psychology	84	127.68	17.82	79
Spanish	61	128.41	17.28	80

[1] All individual scores were transmuted into equivalent scores for 1943. All choices of major subjects were made in the Spring of the students' sophomore year; however students who changed their major during their junior or senior year were placed according to their final choice.

[2] *Norms based on 31,980 freshmen from two hundred forty-one colleges.*[7]

[3] The departments of English Literature and English are combined in this study. Sociology is combined with Economics since for the greater part of the years covered they were included within a single department.

the French, Mathematics, Chemistry, Spanish, and Psychology majors proved to have made significantly higher means (one per cent level) than the students who selected Biology, Political Science, Economics and Sociology, and English. The means of the French, Mathematics, and Chemistry majors were also significantly higher than the History mean (one per cent level), while the Spanish and Psychology means were significantly higher than the History mean at the five per cent level of confidence. The means of the Art and Latin majors were significantly above those of Biology, Political Science, and Economics and Sociology at either the two or five per cent level of confidence. Music majors were found to be neither significantly higher nor lower than any other group of majors at the three levels of confidence employed.

In Table 2 are included 7997 grades of freshman and sophomore students, earned between 1939 and 1947 and separated according to the thirteen divisions of instruction under consideration. As will be observed from the table, the mean grade ranges from 3.72 in Art* to 3.18 in French, Psychology, and Spanish; the standard deviations ranging from 1.17 in French to 0.78 in Biology and English.

The divisions of instruction may be conveniently divided into two groups: the eight 'higher-grading' and the five 'lower-grading.' The higher-grading divisions of Art, Latin, Music, Political Science, English, Economics and Sociology, Biology, and History were all significantly above (one per cent level of confidence) the five divisions of French, Spanish, Psychology, Mathematics, and Chemistry. None of the lower-grading divisions differed significantly from one another, but among the higher-grading groups the means of Art and Latin were significantly above (one per cent level) those of History, Biology, Economics and Sociology, and English. The Art mean grade was also higher than that earned in Political Science (one per cent level) and higher than the Music mean (five per cent level of confidence).

---

* Probably one would expect the Art and Music grades to average higher than those in the other divisions of instruction, since these two are the only 'purely elective' courses among the thirteen divisions. In order to graduate a student must have successfully completed courses in English; History; a foreign language or languages; Chemistry, Mathematics, or Physics; Biology or Psychology; and Economics, Sociology, or Political Science.

It will be observed that the five groups whose means were the lowest on the A. C. E. all selected their major fields of concentration from among the eight higher-grading divisions of instruction. Likewise five of the seven groups who scored relatively high on the A. C. E. subsequently selected their major field from among the five lower-grading divisions of instruction. Only in the case of the Art and Latin departments do we find relatively high-scoring A. C. E. groups choosing their major fields from the higher-grading divisions. The coefficient of correlation by the Product-moment method between mean A. C. E. scores of majors and the mean grades of freshmen and sophomores in the same division of instruction is −.44.

TABLE 2.—MEANS AND STANDARD DEVIATIONS OF FRESHMAN AND SOPHOMORE GRADES FROM 1939 TO 1947 GROUPED ACCORDING TO DIVISION OF INSTRUCTION[1]

Division of Instruction	N	Mean[2]	SD
Art	160	3.72	0.92
Biology	589	3.41	0.78
Chemistry	569	3.25	0.95
Economics & Sociology	450	3.42	0.79
English	1389	3.45	0.78
French	750	3.18	1.17
History	1007	3.40	0.93
Latin	503	3.61	0.90
Mathematics	760	3.23	1.04
Music	379	3.52	0.99
Political Science	328	3.47	0.82
Psychology	341	3.18	0.91
Spanish	772	3.18	1.04

[1] Including only grades of students whose surnames began with A to L and who were in college for two or more years.

[2] Letter grades of A, B, C, D, and F were arbitrarily numbered 5, 4, 3, 2, and 1, respectively.

SUMMARY AND CONCLUSIONS

1) All freshmen in a southern Liberal Arts college for women were given the American Council Examination from 1939 to 1947. Among these students tested there were 1279 who were enrolled in the college during their sophomore year, selected a major field for concentration, and indicated their intention of returning for their junior year.

2) Twelve hundred fifty-two of the sophomores selected one of thirteen divisions of instruction for their major. The freshman A. C. E. scores of this number were transmuted into equivalent scores of the 1943 edition.

3) The seven higher-scoring groups, determined by their means on the A. C. E., were: French, Mathematics, Chemistry, Spanish, Art, Psychology, and Latin. None of these groups differed significantly from one another.

4) The five lower-scoring groups, determined by their A. C. E. means, were: History, English, Economics and Sociology, Political Science, and Biology. No significant differences were obtained among these means.

5) The Music majors, the most variable of the groups, did not differ significantly from any of the other twelve groups.

6) Five of the higher-scoring groups, *i.e.*, French, Mathematics, Chemistry, Psychology, and Spanish majors, earned significantly higher test means than the five lower-scoring groups.

7) Approximately 8000 grades secured from 1939 to 1947, the great majority of which indicated six semester hours of work, were tabulated according to the thirteen divisions of instruction. These were selected by recording the freshman and sophomore grades of all students whose surnames began with A to L inclusive and who attended the college for two or more years.

8) The grades of eight divisions of instruction were significantly above those of the other five, on the average. The five lower-grading divisions were those of French, Mathematics, Chemistry, Psychology, and Spanish.

9) Except for the departments of Art and Music, in which none of the courses are required and to which special talents might be expected to direct students, and Latin, there appears some evidence of an inverse relation between mean A. C. E.

score of majors and mean grade earned by beginning students in the field. The divisions of instruction in which freshman and sophomore students were graded lowest on the average, *i.e.*, Chemistry, Mathematics, Spanish, Psychology, and French, were those that had majors who ranked on the average between the 79th and 86th percentiles on the A. C. E.; whereas the departments that graded their freshmen and sophomores somewhat higher, *i.e.*, Biology, Economics and Sociology, English, History, and Political Science, had majors who on the average ranked between the 68th and 72nd percentiles on the A. C. E. It would appear therefore that choice of major subject bears some relationship to students' A. C. E. scores and to the average grades of those enrolled in the more elementary courses.

10) By combining 'related fields' as Flory did, we find our results to be in agreement with this investigator only insofar as the Social Science group scored lowest of the four divisions. Since the ranking of the other three divisions is at variance with that of Flory, it is quite probable that the agreement in the Social Science division is fortuitous.

REFERENCES

1) C. D. Flory. "The intellectual growth of college students." *J. Educ. Res.*, 1940, 33, 443–451.

2) L. D. Hartson. "Does college training influence test intelligence?" *J. Educ. Psychol.*, 1936, 27, 481–491.

3) A. M. Shuey. "Improvement in scores on the American Council Psychological Examination from freshman to senior year." *J. Educ. Psychol.*, 1948, 39, 417–426.

4) L. L. Thurstone and T. G. Thurstone. "Psychological Examinations, 1940 Norms." *Amer. Counc. Educ. Stud.*, Washington, D. C.

5) ——— ———. "Psychological Examinations, 1941 Norms." *Amer. Counc. Educ. Stud.*, Washington, D. C.

6) ——— ———. "Psychological Examination for College Freshmen 1942 Norms." *Amer. Counc. Educ. Stud.*, Washington, D. C.

7) ——— ———. "Psychological Examination for College Freshmen 1943 Norms." *Amer. Counc. Educ. Stud.*, Washington, D. C.

8) ——— ———. "Psychological Examination for College Freshmen 1944 Norms." *Amer. Counc. Educ. Stud.*, Washington, D. C.

9) ——— ———. "Psychological Examination for College Freshmen 1945 Norms." *Amer. Counc. Educ. Stud.*, Washington, D. C.

10) ——— ———. "Psychological Examination for College Freshmen 1946 Norms." *Amer. Counc. Educ. Stud.*, Washington, D. C.

THE RELATION BETWEEN INTELLIGENCE AND ACHIEVEMENT TEST RESULTS FOR A GROUP OF COMMUNITIES[1]

ROGER T. LENNON
Director, Division of Test Research and Service,
World Book Company

What is the relation between the intelligence level of the elementary-school population of a community and its achievement as measured by standard achievement tests? Is this relation the same from grade to grade or does it tend to increase or decrease with additional schooling? Does this relationship vary from subject to subject and, if so, in what manner? These are questions of considerable theoretical importance and not without practical significance, particularly with reference to such matters as the provision of separate achievement test norms for communities or schools of varying intelligence levels. This paper presents data on the relationship between achievement and intelligence as measured by standard tests, for successive grades and for some important subjects in the elementary school.

SOURCE OF DATA

The data reported were obtained in connection with the standardization of the new (1946) edition of the *Metropolitan Achievement Tests.* They are based on the results for seventy of the two hundred fifty communities which participated in the standardization program; these seventy are communities in which all public school pupils in grades II through VIII (or IX, depending on the organization of the system) were tested, and exclusive of very small communities whose results are based on such few cases as to be of very limited dependability. The communities, however, are for the most part, small ones; about two-thirds are of less than ten thousand population. They are drawn from thirty-six states. The pupils in the various grades took the appropriate battery of both the *Pintner General Ability Test* series, and of the *Metropolitan* series, as listed in Table 1.

---

[1] Paper read at the April, 1949, meeting of the Eastern Psychological Association, Springfield, Massachusetts.

All tests were administered within two months of the opening of schools in the fall of 1946. The figures reported are based on an analysis of a random sample of approximately twenty-five per cent of the total number of cases in each community; for this random sample the mean IQ for each grade in each community was computed, together with the mean standard score in each subtest of the *Metropolitan Achievement Tests* for each grade. It is the relationships among these community averages which are here discussed.

TABLE 1.—BATTERIES OF *Metropolitan Achievement Tests* AND *Pintner General Ability Test* USED IN VARIOUS GRADE LEVELS IN METROPOLITAN STANDARDIZATION PROGRAM

Grade	*Metropolitan Achievement Tests*	*Pintner General Ability Test*
II	Primary I: Word Picture, Word Meaning, Word Recognition, Numbers	Pintner-Cunningham Primary Test, Form C
III	Primary II: Reading, Word Meaning, Arith. Fund., Arith. Prob., Language Usage, Spelling	Pintner-Durost Elementary Test: Scale I (Picture Content), Form B
IV	Elementary: Reading, Vocabulary, Arith. Fund., Arith. Prob., Language Usage, Spelling	Pintner-Durost Elementary Test: Scale II (Reading Content), Form B
V–VI	Intermediate: Reading, Vocabulary, Arith. Fund., Arith. Prob., English, Literature, History, Geography, Science, and Spelling	Pintner General Ability Test: Intermediate Form B, Verbal Series
VII–IX	Advanced: Reading, Vocabulary, Arith. Fund., Arith. Prob., English, Literature, History, Geography, Science, Spelling	Pintner General Ability Test: Intermediate Form B, Verbal Series

The basic data of this paper are summarized in Tables 2 and 3. Table 2 presents the correlations between the average intelligence level (mean IQ) and average achievement level for each grade and subject. In all grades except grade IX these correlations are based on 70 communities; in grade IX they are based on 39 communities. Table 3 presents for each grade the means and standard deviations of the achievement and intelligence test distributions; it is to be noted that these are for distributions of community means, not of individual scores.

SUMMARY OF FINDINGS

Examination of the data presented in Table 2 indicates that all relationships between intelligence and achievement test scores reported are positive; at grade IV and above, with a few exceptions, they are significantly so. (For N = 70, as here, r must equal about .30 for significance at the 1% level.) This is entirely in accord with expectation, although the magnitude of the correlations at the lower grades is perhaps not so great as would have been anticipated. The correlations range from .11 between IQ and what we have called Arithmetic Fundamentals (really Numbers) at the second grade level, to .86 between IQ and Vocabulary score at the eighth grade level; the median value for all correlations is approximately .60. That is to say, in general there is a marked tendency for communities which are above average in IQ to be above average also in achievement, as measured by these tests.

If we turn our attention to the manner in which these relations vary from grade to grade, we note a marked tendency for the correlations to increase as one goes up through the grades; this is clearly revealed by inspection of the right-hand column of Table 2, in which the average value of the correlations for the respective grades is shown. These data indicate that at the second- and third-grade levels, knowledge of the average intelligence test scores for pupils in a community would enable one to improve the prediction of the community's achievement status but slightly, if at all. At the upper grades—VII, VIII, and IX—on the other hand, the tendency is much more pronounced for the communities with high average IQ's to be also the high-achieving communities.

TABLE 2.—CORRELATIONS BETWEEN COMMUNITY MEAN IQ (*Pintner General Ability Test*) AND MEAN STANDARD SCORE ON SUBTESTS OF *Metropolitan Achievement Tests*

Grade	Subject										
	Reading	Word Meaning	Arith. Fund.	Arith. Prob.	Spelling	English	Literature	History-Civics	Geography	Science	Aver.,* all subj.
II	.34[1]	.33	.11[2]								.29
III	.42	.42	.19	.30	.27						.33
IV	.61	.75	.19	.41	.26						.48
V	.68	.75	.25	.44	.37	.58					.54
VI	.72	.81	.13	.43	.62	.60	.63	.50	.67	.62	.60
VII	.79	.85	.42	.50	.54	.66	.74	.63	.70	.67	.67
VIII	.85	.86	.65	.65	.61	.52	.77	.71	.71	.68	.72
IX	.84	.82	.57	.55	.56	.61	.67	.69	.72	.53	.67
Aver.,* all Grades	.66	.74	.36	.48	.49	.60	.70	.64	.70	.63	

[1] Value for both Word Picture and Word Recognition Tests.

[2] Numbers Test.

* Averaged via z transformation.

TABLE 3.—MEANS AND STANDARD DEVIATIONS OF COMMUNITY MEAN STANDARD SCORES, BY GRADE, ON SUBTESTS OF THE *Metropolitan Achievement Tests*

Grade	Subject																					
	Reading		Word Meaning		Arith. Fund.		Arith. Prob.		Spelling		English		Literature		History-Civics		Geography		Science		IQ	
	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ
II	112.5	5.8[1]	121.2	6.2	137.2	3.2[2]															100.7	7.2
III	141.2	5.7	142.6	5.9	137.8	2.5	149.4	2.6	142.7	6.7											100.7	5.5
IV	164.9	7.1	165.7	6.6	154.7	5.1	165.8	4.1	163.7	7.7	164.3	7.1									98.3	6.9
V	180.7	8.1	179.2	7.6	171.6	6.3	178.6	4.9	179.6	6.2	177.9	6.5									100.7	6.0
VI	194.7	7.4	196.0	8.0	195.4	10.4	198.2	6.8	196.6	6.3	195.8	8.2	197.3	7.1	199.1	7.4	197.9	7.6	199.5	8.7	99.5	6.0
VII	204.3	7.5	204.7	7.3	220.2	9.4	211.2	7.2	210.1	6.9	205.5	7.5	206.4	8.3	207.8	7.0	208.5	9.1	208.7	8.5	98.9	5.9
VIII	214.9	7.1	216.0	7.2	242.7	12.3	223.8	9.3	228.6	8.4	222.4	10.0	216.9	8.6	218.0	8.4	216.9	9.9	224.5	10.7	99.5	5.9
IX	221.2	6.4	224.2	6.8	257.4	14.5	233.3	9.4	239.0	9.3	233.2	10.2	222.7	6.8	224.8	7.4	223.2	9.2	234.0	9.1	100	5.5

[1] Word Recognition Test.

[2] Numbers Test

This apparent change in the relationship between upper and lower grades may be to some extent an artifact, reflecting differences in the content of the intelligence tests used. In the second grade, the *Pintner-Cunningham Test* was used, and in the third grade, Scale I, the Picture Content Scale, of the *Pintner-Durost Elementary Test* was used. Both of these are non-reading tests and measure abilities that bear little resemblance to school work; whereas the tests used in grades IV to IX are verbal tests, whose subtests include, among others, arithmetic problems, vocabulary, opposites, and number sequence tests, which are obviously influenced by school experience. Hence, it is not surprising that the intelligence and achievement test results for grades IV to IX should agree more closely than in the earlier grades. However, this would seem to be but part of the story, for even between the fourth and the ninth grades, in which range the same intelligence test was used, there is a steady increase in the degree of correspondence between the two sets of scores. One could argue that the intelligence test results are becoming progressively more dependent on, more a reflection of the influences of, schooling; but there is nothing in the data to show in which direction causative influences, if any, are operating.

A second hypothesis would be that the increasing correlation between average IQ and average achievement with increasing schooling is in part a function of greater variability in either measure as one goes up through the grades. The data in Table 3 lend some support to this belief. The variability of community mean IQ's remains quite constant throughout the grades, to be sure, but there is a clear tendency for the communities to become more variable in achievement at the higher grades, as examination of the standard deviations reveals.

What of the relation of community mean IQ to community performance in the separate subjects? For communities, as for individuals, there are marked differences among subjects at all grade levels. In general, the greatest correlation tends to be between the reading measures and the intelligence measures; at the seventh-, eighth- and ninth-grade levels, these correlations almost approach the magnitude that one would expect to find between two measures of intelligence or between two reading tests. Thus, knowing the average intelligence level of pupils in these grades, as measured by the tests used, one could estimate with considerable accuracy their average reading achievement, or vice versa. The correlations tend to be smallest between IQ and what we may term the 'skill' subjects, namely, arithmetic and spelling; for Arithmetic Fundamentals in grades II to VI, knowledge of intelligence is of little help in predicting achievement. At the lower grades the correlations between Arithmetic Fundamentals and IQ are significantly smaller than the correlations between Arithmetic Problem Solving and intelligence, but this is not the case in grades VII to IX. Between the reading tests, on the one hand, and the 'skill' tests on the other, are the so-called 'content' subjects—Literature, History and Civics, Geography and Science; the correlations between community means in these subjects and intelligence range from .50 to .77.

The attempt to relate achievement measures to intelligence test results has been of persistent concern to test-makers and test-users—and legitimately so, for the desire to evaluate an individual's performance in terms of his 'capacity' is a praiseworthy one. Accomplishment quotient techniques and cognate devices represent one approach to this problem, albeit scarcely a satisfactory one. The provision of separate norms for the interpretation of scores for individual pupils of varying IQ levels, as is done for certain achievement tests, represents another approach. Some writers have suggested that it would be desirable to provide achievement test norms for school systems of varying intelligence levels on the ground that such norms would permit the school administrator to compare achievement in his system with achievement in systems whose pupils are similar in intelligence. This suggestion assumes that variation in the average intelligence level is associated with variation in achievement to a degree sufficient to make it worth while for practical purposes to establish differentiated norms; as mentioned in the opening sentences of this article, the data herein reported are relevant to this question.

The data presented in Table 2 indicate, in the writer's judgment, that in grades II through V, at least, the relationships between the intelligence and achievement levels of a community, with the single exception of those for reading, are not sufficiently large to warrant the establishment of differentiated norms for school systems of varying average intelligence levels. In grades VI through IX the correlations, with the exception of those for arithmetic, spelling and English, are of such magnitude that one may expect communities scoring high in intelligence tests also to do well on achievement tests with considerable consistency. If the magnitude of this relationship were the only factor to be considered in determining the desirability of differentiated norms, one could argue in favor of such norms at the upper grades except for the subjects named.

There are, however, other elements to be considered in deciding, both on practical and theoretical grounds, whether separate norms of this character are desirable. From a practical standpoint, the provision of multiple sets of norms is expensive, both in terms of the cost of obtaining adequate data and in terms of making the additional normative data available; and it is doubtful that there is sufficient sophistication on the part of the ordinary test-user to make it likely that such differentiated norms would be generally understood and properly used. On theoretical grounds, we might cite the problem of the validity of most currently available intelligence tests as measures of 'capacity' in any real sense, and the corollary problem of the community of function measured by present intelligence and achievement tests. Even if the practical problems in the construction and use of differentiated norms could be coped with successfully, the theoretical justification for such norms seems, to the writer, dubious to say the least.

A STUDY OF HIGH-SCHOOL ACADEMIC INDICES AS A CRITERION FOR COLLEGE ADMISSION

ISMAEL RODRÍGUEZ BOU and FRANKLIN L. STOVALL
Superior Educational Council
University of Puerto Rico
Rio Piedras, Puerto Rico

INTRODUCTION

One of the many problems with which the University of Puerto Rico is faced is that of selecting wisely and well its student body from among the numerous individuals who desire to enroll. This problem exists even now and, without a doubt, will become more serious in the future. At present, the University and other institutions of higher learning on the Island are not able to accommodate all persons who desire to matriculate. Some kind of a screening process seems necessary in order that these institutions shall be able to admit those people who can profit to the greatest extent from college training.

In such a screening process many factors must be taken into consideration in order that those individuals who must be screened out or eliminated (if such be the case) will be the individuals who are least likely to succeed at the college level. One of the factors which has carried most weight in the selection of students by institutions of higher learning has been and still is the factor of high-school grades or previous academic record. The University of Puerto Rico prior to the school year 1946–47 used this factor as a sole criterion for selection.

How reliable is such a criterion? To what degree of accuracy is it possible to predict college grades from high-school grades? Are letter grades of 'A,' 'B,' 'C,' etc. from the various high schools of Puerto Rico equivalent? Do those individuals with the same letter grade averages in high school possess approximately the same degree or amount of preparation for college work?

In order to obtain more accurate and reliable answers to these and similar questions it was decided to make a comprehensive and scientific study of the problem. Accordingly, a study was sponsored by the Superior Educational Council of the University of Puerto Rico with Dr. Ismael Rodríguez Bou, Permanent

Secretary, as Director. This paper is in the form of a summary report* of the study—the purpose, methods and technique, the presentation and analysis of data, the conclusions and, finally, the recommendations.

PURPOSE

The purpose of this study was two-fold:

1) To evaluate the high-school academic record as a criterion for admission to college. With what degree of accuracy is it possible to predict grades in the University of Puerto Rico from grades made in high school?

2) To determine the relative importance of letter grade averages as given to students by the teachers of the various high schools of the Island. Is an 'A' or 'B' grade in one high school equivalent to an 'A' or 'B' grade in some other high school? Are students from different high schools with the same letter grade averages equally prepared for academic work in college?

METHODS AND TECHNIQUES

In the initial stages of this study, the grades made in the University by eight hundred fifty-three freshmen just out of high school (fifty-seven different schools) who were enrolled in the University during the year 1945–46 were obtained along with their respective grades in high school. The University grades made by each student in all courses completed during the 1945–46 school year were averaged in order to get an average grade for each of the eight hundred fifty-three freshmen included in the study. Similarly, an average of high-school grades was computed for each student. In computing the University and high-school averages an 'A' was given the value of 4, a 'B' the value of 3, a 'C' the value of 2, a 'D' the value of 1, and an 'F' a value of 0. The lowest high-school grade point average of any student was 2, for students with less than a 'C' average in high school are not admitted to the University (except veterans who pass a special examination).

These averages were then compared and a coefficient of correlation computed. A regression equation was worked out by means of which University grade point averages were predicted from high-school grade point averages. The predicted grades were then compared directly with the grades actually made in the University and the extreme variations noted. An attempt was made to relate these extreme cases to particular high schools.

---

* A complete comprehensive detailed report of this study has been published as English and Spanish Editions by the Superior Educational Council of Puerto Rico. Copies are available upon request.

Some months later the study was extended to include an additional group of one thousand one hundred fifty-seven students who had matriculated in the University for the school year 1944–45 and who had come directly from the high schools of the Island. The same methods and techniques employed in an analysis of the grades of the eight hundred fifty-three students were used in dealing with this last group of one thousand one hundred fifty-seven students.

Also, it was possible still later to study in the same manner two hundred forty-eight and two hundred thirty-four students who entered in 1944–45 and 1945–46, respectively, the College of Agriculture and Mechanical Arts of the University of Puerto Rico, located at Mayagüez. Then one hundred eight students who were freshmen at Polytechnic Institute in 1945–46 were included. Finally, the combined total of two thousand six hundred students was studied.*

DATA FROM 2,600 CASES

Averages of high-school academic indices and University grade points for the several class groups (a total of 2,600 cases) are given in Table I along with the standard deviations and coefficients of correlation between high-school indices and grades made in college. Since high-school graduates who have less than a 'C' average in all their high-school work are not admitted to the University of Puerto Rico and to Polytechnic Institute, the high-school academic indices for the above groups range from 2.00 to 4. In effect, then, the students admitted to college in Puerto Rico come from the upper or 'right' half of at least approximately a normal distribution representing the total of all high-school graduates for the years involved. Grades made in college, however, range from 0 to 4 and follow a normal curve of distribution.

---

* The data here deal only with the combined total of all cases and represent graduates of seventy-three different high schools in the Island, both public and private. The total, however, includes six students at Polytechnic Institute who graduated from high school outside of Puerto Rico.

Some interesting comparisons are possible by an examination of Table I. Columns 2 and 3 give the means or arithmetic averages of high-school and college indices. The mean high-school index for students enrolling in the University of Puerto Rico (2,010 cases) is 2.71, for the College of Agriculture at Mayagüez, 2.47 (482 cases), and for Polytechnic Institute at San Germán, 2.93 (only 108 cases for one year). The average high-school index for all 2,600 cases (including six students at the Institute

TABLE I.—MEANS, SIGMAS, AND COEFFICIENTS OF CORRELATION COMPUTED FOR THE SEVERAL GROUPS

Groups	Means		Sigmas		
	High School	University Grades	High School	University Grades	r's
University (853)	2.74	2.20	0.52	0.67	0.57
University (1157)	2.69	1.97	0.53	0.78	0.51
Total: 2010	2.71	2.07	0.53	0.74	0.54
College of Mayagüez (248)	2.48	2.18	0.54	0.60	0.52
College of Mayagüez (234)	2.45	2.01	0.59	0.79	0.70
Total: 482	2.47	2.10	0.57	0.70	0.62
Polytechnic Institute (108)*	2.93	2.21	0.51	0.50	0.54
Grand Total: 2600	2.67	2.08	0.53	0.73	0.57

* Includes six students who graduated from high schools outside of Puerto Rico.

who graduated from high school outside of Puerto Rico) is 2.67, or a low 'B' average. The mean college index for the University is 2.07; for the College of Agriculture, 2.10; for the Institute, 2.21; and for all combined, 2.08. In columns 4 and 5 of Table I are given the corresponding standard deviations.

In the last column of Table I the coefficients of correlation between high-school and college indices are listed. The averages of these coefficients were computed by squaring each coefficient, securing an average of these squares, and extracting the square root. The correlation for the University and the Institute are the same, .54. The correlation for the College of Agriculture is the highest, .62. The coefficient of correlation for all 2,600 cases combined into one large sample is .57. All of these coefficients are positive and satistically significant, but they are quite low for purposes of prediction. When regression equations were worked out, the error of prediction was so large that for practical purposes it is difficult to predict accurately college grades from a knowledge of the high-school grades. This means then that high-school indices as a criterion for the selection of college students is not very reliable, especially in regard to students with a low 'C,' 'C,' or high 'C' average.

Table II lists the seventy-three public and private high schools which constituted the source for the student groups studied. The total number of students and the average high-school index for all the students from each particular school are given in columns 2 and 3. In column 4 the actual grade point averages earned by students during their first year of college are listed. The grade point averages in Column 5 were predicted from a regression equation. For the individual, the error of estimate is so high that knowledge of the student's high-school index does not enable one to predict with any degree of accuracy the degree of success he may attain in college. The differences in grades actually obtained and grades predicted by means of the regression equation are shown in the last column of the table. The range of these differences is all the way from +1.43 (one student from a small private school) to −0.84 (one student from a small public school). A study of this column of differences reveals that students from the largest high schools and a majority of the private schools on the whole make better grades in college than predicted for them on the basis of the regression equation.

TABLE II.—A COMPARISON OF THE ACTUAL AND PREDICTED UNIVERSITY GRADES BY TOWNS

Town	No. of Students	H-S Index	Grades in U.	Predicted Grades	Difference
1. Adjuntas	11	2.92	2.21	2.26	−0.05
2. Aguada	6	3.14	1.64	2.42	−0.78
3. Aguadilla	75	2.68	2.19	2.08	+0.11
4. Aibonito	20	2.81	1.86	2.16	−0.30
5. Arecibo	99	2.59	2.00	2.02	−0.02
6. Bayamón	128	2.67	1.95	2.05	−0.10
7. Cabo Rojo	39	2.50	1.96	1.98	−0.02
8. Caguas	149	2.72	2.17	2.13	+0.04
9. Canóvanas	18	2.61	1.41	1.96	−0.55
10. Carolina	71	2.71	1.81	2.09	−0.28
11. Cayey	52	2.65	2.25	2.04	+0.21
12. Ciales	11	2.76	2.33	2.13	+0.20
13. Coamo	41	2.93	1.87	2.28	−0.41
14. Comerío	16	2.87	1.92	2.18	−0.26
15. Corozal	13	3.09	1.96	2.36	−0.40
16. Fajardo	50	2.57	1.94	1.97	−0.03
17. Guánica	10	2.45	1.62	1.86	−0.24
18. Guayama	62	2.70	1.93	2.07	−0.14
19. Hormigueros	7	2.74	1.59	2.19	−0.60
20. Humacao	43	2.71	2.02	2.09	−0.07
21. Isabela	23	2.69	1.92	2.12	−0.20
22. Jayuya	10	2.54	1.82	1.99	−0.17
23. Juana Díaz	1	2.00	1.18	1.91	−0.73
24. Juncos	15	2.53	1.91	1.95	−0.04
25. Lares	36	2.65	1.97	2.02	−0.05
26. Manatí	69	2.82	1.96	2.17	−0.21
27. Mayagüez	197	2.42	2.04	1.99	+0.05
28. Orocovis	1	3.39	2.04	2.88	−0.84
29. Ponce	166	2.66	2.22	2.07	+0.15
30. Río Grande	24	2.83	1.58	2.17	−0.59
31. Río Piedras (U.P.R.)	104	2.70	2.39	1.99	+0.40
32. R. Piedras High School	151	2.60	1.97	2.01	−0.04
33. R. Piedras (Montalvo)	18	2.40	1.57	1.82	−0.25
34. Salinas	27	2.77	1.87	2.14	−0.27
35. San Germán	37	2.62	2.18	2.11	+0.07
36. S. Juan (Superior Central)	412	2.68	2.25	2.08	+0.17
37. S. Juan (Superior Nocturna)	9	2.72	2.39	2.04	+0.35
38. S. Juan (Negoc. Adultos)	4	2.81	2.37	2.06	+0.31
39. San Sebastián	33	2.76	1.86	2.12	−0.26
40. Toa Baja	10	2.43	1.77	1.95	−0.18
41. Utúado	48	2.70	1.87	2.10	−0.23
42. Vega Baja	37	2.77	1.95	2.22	−0.27

TABLE II.—(*Continued*)

Town	No. of Students	H-S Index	Grades in U.	Predicted Grades	Difference
43. Vieques	16	2.38	1.56	1.87	−0.31
44. Yabucoa	16	2.70	1.98	2.04	−0.06
45. Yauco	52	2.63	2.07	2.11	−0.04
Private Schools					
46. St. Joseph (Aibonito)	1	2.37	1.30	1.73	−0.43
47. San Felipe (Arecibo)	2	2.87	1.98	2.13	−0.15
48. Ac. Bautista (Barranquitas)	9	2.66	1.79	2.64	−0.85
49. Mons. Willinger (Barranquitas)	17	2.83	1.95	2.15	−0.20
50. Santa Rosa (Bayamón)	7	2.81	1.52	2.25	−0.73
51. Presbyterian (Lajas)	2	3.57	2.03	2.55	−0.52
52. San Luis (Lajas)	11	2.74	1.55	2.08	−0.53
53. Inmaculada (Manatí)	1	2.45	2.00	1.79	+0.21
54. Inmaculada (Mayagüez)	12	2.81	2.65	2.30	+0.35
55. Santa Teresa (Naranjito)	5	2.45	2.10	2.13	−0.03
56. Col. Ponceño de Varones	12	2.67	2.23	2.15	+0.08
57. Liceo Ponceño	4	2.84	2.71	2.21	+0.50
58. Modern Business College	2	2.04	1.83	1.48	+0.35
59. La Milagrosa (R. Piedras)	5	2.37	1.79	1.81	−0.02
60. San Antonio (R. Piedras)	9	2.53	1.69	1.89	−0.20
61. San Jorge (R. Piedras)	1	2.31	1.71	1.88	−0.17
62. San José (R. Piedras)	7	2.00	1.64	1.93	−0.29
63. P.R. High School of Commerce (Río Piedras)	1	2.03	2.90	1.47	+1.43
64. San José (San Germán)	5	3.47	2.51	2.50	+0.01
65. Acad. Católica (S. Juan)	5	3.06	2.18	2.32	−0.14
66. La Inmaculada (Santurce)	5	2.91	2.22	2.18	+0.04
67. Col. Puertorriqueño (Santurce)	12	2.77	2.56	2.08	+0.48
68. Perpetuo Socorro (Santurce)	10	3.31	2.36	2.57	−0.21
69. Sagrado Corazón (Santurce)	2	2.39	2.52	1.84	+0.68
70. Blanche Kellogg (Santurce)	3	2.57	2.34	1.94	+0.40
71. Santa Rita (S. Sebastián)	7	3.01	1.87	2.25	−0.38
72. Holy Rosary (Yabucoa)	2	2.73	2.12	2.19	−0.07
73. Holy Rosary (Yauco)	2	2.39	2.28	2.13	+0.15

Students from the smaller high schools do not attain the level expected of them.

This fact is more forcefully presented in Table III where nine of the larger and twenty-three of the smaller public high schools are listed and tabulated on the basis of the weighted differences between grades actually obtained in college and grades predicted. In column 1 intervals in terms of 'difference' units are set up. The number of students and the actual weighted 'differences' are

TABLE III.—DISTRIBUTION OF THE DIFFERENCES BETWEEN COLLEGE GRADES OBTAINED AND COLLEGE GRADES PREDICTED FOR NINE LARGE AND TWENTY-THREE SMALL HIGH SCHOOLS

Intervals	High School	N	Difference
0.20– 0.29	Ciales	11	0.20
0.10– 0.19	San Juan*	412	0.17
	Ponce*	166	0.15
0.00– 0.09	Mayagüez*	197	0.05
	Caguas*	149	0.04
−0.10−−0.01	Arecibo*	99	−0.02
	Cabo Rojo	39	−0.02
	Juncos	15	−0.04
	Río Piedras*	151	−0.04
	Yauco	52	−0.04
	Adjūntas	11	−0.05
	Lares	36	−0.05
	Yabucoa	16	−0.06
	Humacao*	43	−0.07
	Bayamón*	128	−0.10
−0.20−−0.11	Guayama*	62	−0.14
	Jayuya	10	−0.17
	Toa Baja	16	−0.18
	Isabela	23	−0.20
−0.30−−0.21	Utuado	48	−0.23
	Guánica	10	−0.24
	San Sebastián	33	−0.26
	Comerío	16	−0.26
	Salinas	27	−0.27
	Aibonito	20	−0.30
−0.40−−0.31	Vieques	16	−0.31
	Corozal	13	−0.40
−0.50−−0.41	Coamo	41	−0.41
−0.60−−0.51	Canóvanas	18	−0.55
	Río Grande	24	−0.59
	Hormigueros	7	−0.60
−0.70−−0.61			
−0.80−−0.71	Aguada	6	−0.78

* High schools with enrollments of 500 or more.

given in the last two columns. An asterisk has been placed by the name of each high school with a total enrollment of five hundred or over. The other twenty-three high schools included in the table have enrollments under three hundred.

It will be noticed that the large high schools are high in the distribution and that the small schools are low. This means that in general the students from the large high schools made indices in the colleges as good as, or better than, those predicted for them, but that the students from most of the small high schools actually earned college indices lower than those predicted for them on the basis of their high-school grades. Students with the same letter grades from small and large high schools do not have, it seems, the same degree of training and preparation for college work. High-school indices from various high schools and based on the system of letter grades in use do not appear to be equivalent.

CONCLUSIONS

On the basis of an analysis of the data used in this study, the following conclusions seem justified:

1) The high-school academic index is not very reliable as a criterion for selecting college students.

2) There is a positive relationship between high-school and college indices but the degree of correlation is so low that high-school indices cannot be used for predicting accurately college indices. The errors of estimate are high.

3) Students applying for admission to college who have the same high-school indices do not necessarily possess the same degree of training and preparation for college work.

4) An 'A' or 'B' grade in one high school is not necessarily equivalent to the corresponding 'A' or 'B' grade in another high school.

5) Students in general from the large high schools (enrollment of five hundred or more) make higher indices in college than do students from small high schools (enrollment of three hundred or less) with the same high-school indices.

6) The chances for success in college based on academic record are greater for students from the large high schools than for students from the small high schools with the same high-school indices.

7) Students with a 'C' average from large high schools make on the whole higher grades in college than do 'B' students from small schools.

8) The indications are that a grave injustice may be done students from large high schools with a high-school index slightly below that required for admission to college. They may be in many instances better prepared for college work than are students from small high schools who are admitted to college because they have high-school indices a few points higher.

RECOMMENDATIONS

The purposes of this study have been realized. The data have been collected, systematized, analyzed, and interpreted. Conclusions which seem to be well justified by the objective techniques employed have been listed. This paper should not end here, however. Certain suggestions of a constructive nature are in order. These recommendations concern the general problem of providing some kind of a screening process whereby those selected for college training shall be the individuals from among all the applicants for admission who may profit most from college work. Such a screening process would result in the admission of those students who have the greatest chance of succeeding at the college level. The recommendations offered here concerning this general problem may be conveniently classified under three main heads: (1) Those designed to improve the reliability of high school indices as a criterion for selection (1–5 inclusive); (2) those designed to utilize other factors as criteria for college entrance (6 and 7); and, finally, (3) those designed to extend and expand the educational opportunities and facilities of the Island above the high-school level (8–10 inclusive). Such extension and expansion would tend to change the nature of the selection process itself.

1) Improve instruction in all the public schools of the Island. This involves the provisions for more adequate school facilities and superior teachers.

2) Improve instruction, especially in the small high schools in order that graduates of these schools may be equally prepared (as graduates from large high schools) for college work.

3) A more uniform and standardized system of assigning school marks might be worked out.

4) Consolidate, in so far as is feasible, the small high schools of the Island. Distances within the Island are short and children may be readily transported by bus. Such consolidations would serve to place the advantages of large high schools within reach of all.

5) The Department of Education and the University of Puerto Rico should collaborate in establishing a division of Measurements and Guidance. Such a division would work throughout the year in the preparation of tests and other measuring instruments and in the development of techniques for student guidance. A comprehensive testing program would be set up for the public schools of the entire Island. The test results would be cumulative for each child and would be valuable information which might be used for a variety of purposes. Among other things the test results could be used as another criterion for admission to college and for the purpose of assisting students in choosing the courses in college from which they might derive the greatest benefit. If this were possible, the money which may be spent in giving students instruction in the lines of study for which they are not suited, could be used to greater advantage.

It seems reasonable to assume that the very high mortality rate which now exists might be considerably reduced by a good guidance program. From the economic point of view we should endeavor to hold students at the college level until they finish their studies. The per capita expenditure at the college level is very great. Moreover, mortality at this level involves individuals of high achievement and great capacity for learning. This in itself presents a problem of paramount importance since our country needs people with more training and ability in different fields of education to help in the solution of its social and economic problems.

6) High-school indices should not be used as a sole criterion for selection of college students.

7) Standardized objective tests of various kinds should be utilized as a screening device. Comprehensive tests designed to measure academic aptitude, reading proficiency (vocabulary and comprehension), achievement, and certain aspects of personality development would yield information of great value for all concerned. Such test results are essential for educational, vocational, and personal guidance. Many such tests are already

available and new ones could be constructed to meet the specific needs of Puerto Ricans.

8) Extend and expand the facilities of the University of Puerto Rico and all of its branches. The University should be prepared to meet the needs of a student body of from 15,000 to 20,000 within a few years. As a senior university and graduate school, it should be on a par with the best of them.

9) A system of public junior colleges offering various types of two-year terminal courses should be established. The Island might be divided into four or five regions with a public junior college located in each and accessible to the high-school graduates of that region.

10) Public business, technical, and vocational schools of various kinds to meet the needs of the people of the Island might be provided after a study is made to determine the needs of the people along these lines.

In conclusion, it should be stated with considerable emphasis, that additional research on this and related problems is sorely needed.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 41 October, 1950 Number 6

Educational Psychology for Teachers

INTRODUCTION

In 1947, the Executive Committee of the National Society of College Teachers of Education organized several committees to study the various aspects of teacher education. One of these committees was assigned the task of recommending the content, sequence, organization and teaching procedure for courses in educational psychology.[1] This committee has made three reports of progress. The first report, made in 1948, consisted of seven papers outlining the facts and principles from several areas of psychology as they relate to curriculum development.[2] The second report, made in 1949, consisted of five papers dealing with the nature of educational psychology and the contributions it has made and can make to the preparation of teachers.[3] The third report consists of the six papers presented here. They *deal with the content of educational psychology in a limited* number of areas.[4]

---

[1] The Chairman of this committee is Walter W. Cook, University of Minnesota; the vice-chairman is Wm. Clark Trow, University of Michigan; the sponsor for the Executive Committee of the Society is Frank S. Freeman, Cornell University.

[2] "The Psychological Basis of the Modern Curriculum," *The Journal of Educational Psychology*, March 1948. Vol. 39, pp. 129–169.

[3] "Educational Psychology in the Education of Teachers," *The Journal of Educational Psychology*, May 1949. Vol. 40, pp. 257–294.

[4] The six papers presented here—"Psychology of Group Behavior," "What Educational Measurement in the Education of Teachers," "How Can the Psychology of Development in Infancy and Childhood Help Teachers," "What Teachers Should Know about the Psychology of Adolescence," "What the Psychology of Learning Has To Contribute to the Education of the Teacher" and "The Study of Individual Differences in the Education of Teachers"—were presented at the Atlantic City Meeting of the National Society of College Teachers of Education, February 27 and 28, 1950.

1) PSYCHOLOGY OF GROUP BEHAVIOR: THE CLASS AS A GROUP

WM. CLARK TROW, ALVIN E. ZANDER, WM. C. MORSE
and DAVID H. JENKINS
School of Education, University of Michigan*

Social psychology has been experiencing a marked development in recent years; and because of the many implications for learning situations, those tilling the educational fields should be alert to the new points of view and new findings which are emerging. This statement does not imply that individual educational psychology is to be discarded, but rather that it is now directly complemented by the basic socio-psychological concept of the group and the consideration of intra-group relationships. As long as sociologists confined their attention largely to such social groupings as crowds and mobs, criminals and delinquents, the family, and to census groups with racial and nationality characteristics, the help they could furnish to the classroom teacher was relatively slight. But with the development of field theory and the study of interaction of individuals in a face-to-face group, and more specifically with the coming of the Iowa studies of democratic, autocratic and *laissez-faire* leadership, followed by the energetic labors of those in the field of group dynamics, the picture has changed. To this has been added the later Freudian influence in the mental hygiene movement, its expansion in the area of inter-personal relationships, and the exploitation of such treatment techniques as those of group work and play therapy. We are forced to ask ourselves whether the school class is a group, and, if it is, what this should mean to educational psychologists whose task it is to introduce teachers to the principles which should aid them in developing the best possible environment for learning in their classrooms.

---

* Dr. Trow is Professor of Educational Psychology; Dr. Zander is Assistant Professor of Educational Psychology and Program Director of the Research Center for Group Dynamics; Dr. Morse is Assistant Professor of Educational Psychology and Director of the University Fresh Air Camp; and Mr. Jenkins is Lecturer in the School of Education and Field Consultant in Community Adult Education.

DEVELOPMENTAL BACKGROUND

It should be recognized at the outset that educational psychology has from the beginning devoted itself almost exclusively to modifying the responses of individuals to more or less separate stimuli. The principles of learning, derived from the performances of laboratory animals and sometimes of children, though the results were brought together statistically, have been applied to the individual learner; and his performance has been tested by presenting him with a series of tasks to perform, and measuring his success in performing them. To describe the educational psychology of the past and the present in this way is not to belittle it. Tremendous improvements have been made in instructional materials and methods as a consequence of this view. We can well feel proud of the contributions of our colleagues and wish for their continuance, for there is much more to be done. After all, individuals are individuals, and they are probably here to stay!

The single-line, teacher-pupil relationship, however, has other sources than the psychological laboratory. There seem to have been changing patterns in our educational assumptions as to the most effective and desirable learning situations for the pupil. At one time the tutorial arrangement, the scholar and the single student in a face-to-face relationship, was felt to be most nearly ideal. And it may be for certain kinds of learning. But the practical situation in our public schools has not, of course, permitted this kind of teacher-pupil ratio; so we tried to make our classes of twenty-five or more pupils into twenty-five simultaneous one-to-one relationships. At any rate we followed this pattern, in our classwork, of teacher control, assignment, and class discussion, all dependent on the teacher-pupil-teacher-pupil kind of interaction. In this tradition we not only have emphasized the importance of the individual pupil of the subject-centered curriculum, but also of individualized instruction, and the child-centered school.

This arrangement tended to be strengthened by virtue of the fact that it provided a more direct system of control. Any break in the line, with consequent spontaneous interaction among pupils might well mean that the teacher had lost that control which he felt it necessary to maintain. If the class were allowed to become an interacting group, the behavior of the pupils would presumably not be contributing to the learning goals which the teacher had in mind. Thus, 'groupiness' implied 'bad discipline.'

Two factors have probably contributed to the movement away from this tutorial conception of our classrooms: the increasing interest and attention being given to social learning, and the awareness that the classrooms are, potentially at least, social situations. With the acceptance of the broader social goals of learning, no longer restricted to scholarly and intellectual activities alone, dependence on the tutorial tradition began to lessen, and the potentialities of the class as a medium for instruction in social learning became clearer.

The point where modern social psychology can offer desirable additions to the individualized approach lies in a recognition of the complex nature of what has in the past been rather loosely referred to as the stimulus situation when this situation is largely made up of other persons. The exploration of this phenomenon, and of the function of perceptual and conceptual processes in relation to it, is the chief contribution of the gestalt psychologists, whose point of view the late Kurt Lewin was largely responsible for bringing over into the interaction field of social psychology. Teachers have long known that pupils responded to other stimuli than the words of wisdom emanating from behind the teacher's desk. But the teacher's task was to eliminate such distraction so far as possible. And while this is still often desirable, we are now interested in these other stimuli also, in the interactions of the pupils among themselves and with the teacher. We are asking, what are the implications of viewing the class not merely as a number of individuals in the same room, but as a group?

The exposition of this point of view in education did not have to wait for the recent developments in social psychology. Although the tone is definitely authoritarian, beginnings are found, for example in a volume entitled *School Discipline*,[1] by William C. Bagley, published in 1917. In this volume Bagley discussed in some detail the problem of what he called the "unruly school." He pointed out twin antithetical causes: "harsh and unsympathetic treatment," and "indulgence and weakness of control," conditions not too far removed from frustrating autocracy and *laissez-faire*, respectively. He went on to indicate some of the "difficulties of reconciling the opposing ideals of individualism and collectivism." For transforming the unruly school he included among other conditions, "the importance of the objective attitude, and stimulating group responsibility."

Likewise many school practices, particularly in the extracurricular field, have laid a foundation for group interaction. For a number of years group games and sports provided for coöperative as well as for individual effort, and teacher-sponsored 'activities' of the hobby-club variety tended to promote more informal teacher-pupil relationships. The project method, while it chiefly emphasized individual performance, also had a place for group activities. With the activity program came the educational heyday of group participation involving the imitation of adult activities in stores, post offices and the like, but largely employed as a means for motivating learning and providing practice in the traditional subjects.

However, in nearly if not all of these situations, the teacher is set off against the class. His view of the class as a kind of unit is exemplified when the teacher asks a question and then says "Class," calling for all to respond more or less in unison. The teacher is boss, though at times he would tolerate some freedom of action on the part of the children that would permit some release of tensions. Even when an 'audience situation' is provided for pupils to read or recite passages they had learned, the same condition maintains. Similarly in matters of student deportment, now usually referred to as citizenship, the teacher is the interpreter of the mores of the culture for the pupils, and serves as judge, jury, and lord high executioner, all bundled into one.

In some schools, the system of student government, with a student council, ideally shifts some of the responsibility to the pupils and permits pupil interaction and group decisions. Similarly, in what is referred to as teacher-pupil planning the teacher forsakes his antithetical position and becomes an actual group member in the rôle of a resource person. It becomes clear that there has been a long period of gradual change in theory which has been followed by practice in some schools, the majority however probably trailing far behind. At any rate, it may be concluded that education is ready for a systematic overhauling of its theory and practice in dealing with the class as a group, and that it is the proper task of educational psychologists to lead the way.

CONCLUSIONS FROM RESEARCH IN GROUP DYNAMICS

First, in order to explore some of the possible directions that our inquiry might take, let us review briefly a few of the research findings that deal with group functioning and group interrelations in a wide variety of social settings. Although teachers work with groups and are daily troubled or aided by group phenomena in their classrooms, there has been strikingly little research on the dynamics of classroom groups. It is often difficult to identify and study the many forces at work in a classroom situation, but recent research in group dynamics indicates that it is possible to develop the necessary theoretical formulations, hypotheses, and measuring methods for testing these hypotheses. The task remains to identify those areas in which we feel the presence of group phenomena is most relevant to the classroom setting. We have much to learn about the forces involved in the relationships among students, and between students and teacher. Since the relationship between teacher and class-groups, for example, is by its very nature changing and flexible, it is important that the concepts employed be adequate to deal with the dynamics of relationships, involving changing relationships among persons, and changing perceptions of the teacher and the class, as the members acquire new insights and learnings.

A number of assertions from recent research in group dynamics have both theoretical and practical value for the field of educational psychology and teaching methods. This list is not exhaustive and there will be no attempt to describe the nature of the studies from which these data are derived. Many of these findings are from laboratory investigations with groups, but a sufficient number of them were obtained in field-experiment settings to indicate that work of this nature can readily be done in the actual classroom setting, as well as in the laboratory. Some of these assertions are well-tested and validated. Others are less well proven. All of them have relevance and promise for educational psychology.

1) The attitudes of an individual have their anchorage in the groups to which he belongs. Present evidence makes it apparent

that many attitudes can be changed more easily by making changes in certain properties of the group than by directly teaching the individuals, as individuals, even in a classroom audience situation.[9,10]

2) The conduct and beliefs of pupils is regulated in large measure by the small groups within a classroom, such as friendship cliques, and the cohesive groups of students within a school. These groups demand conformity from their members to certain group standards, and the more cohesive the group, the greater is its power over the member.[2,5]

3) In some instances failure to learn may be advantageously conceptualized as resistance to change, using resistance here in the same sense as the therapist uses it in his relationships with a patient.* For example, the group standards developed by persons who were learning a motor task quite similar to a previously perfected one, and who were simply told what they were to do, were entirely different from the group standards developed in a group in which the learners participated in a discussion and made group decisions about the necessity for, and the nature of, the new task to be learned. Those who participated in the discussion learned much more, more rapidly, and with much less aggression and resentment toward the persons inducing them to make this change.[2,17]

4) When frustrations are met, highly cohesive groups maintain their effort in movement toward the group goal much more vigorously and effectively than do groups of low cohesiveness.[6]

5) *Groups, especially those similar to classroom groups, can* be disrupted into separate cliques; or this threat of disruption can be eliminated, by the alteration of forces which determine the attractiveness of the group for the members. (For example, helping them to become aware of the strength of attraction they have for each other, or the degree to which membership in the group provides a way to achieve things they value highly.) This condition can be brought about most easily when the members become aware of the forces influencing them, but it can also be effected by an outsider, such as a teacher, who adroitly helps

---

* It should be noted, however, that failure to change may be due to such '*resistance.*' *There may be an inadequate set, unsatisfactory motivation*, inability to comply with the demands of the goal or a rational non-acceptance of a new position.

the group to change the impact and strength of these forces surrounding and within their group.[16]

6) The training of persons for effective social action such as performance in school or civic service, can lead to greater effectiveness of effort by the trainees if they are members of a group which is being trained to work as a group, than will result if they are merely individuals in an audience situation.[11]

7) The amount of interaction among students in a class is determined in part by group factors. For example, in highly cohesive groups arriving at a decision that has general approval, the person whose viewpoint is too different from that of the rest will be rejected—that is, ignored. In a less well knit group, in which the discussion is not directed to a group decision, the deviate member is likely to get more comments directed to him than the person whose ideas are quite similar to those of the rest of the group.[15]

8) When the members see themselves competing for their own individual goals which make coöperative effort impossible, there is disruption of the ready communication of ideas, the coördination of efforts and the friendliness and pride in one's group which are basic to class harmony and effectiveness. The competitive grading system commonly used today is an illustration in that it creates mutually exclusive goals among the members of a class group.[3,4]

9) The group climate or style of group life can have an important influence on the member's personalities. One such style of group life can develop hostile, obedient, uncreative, 'goldbrickers'; another can produce confused, purposeless, competitive, drifters; and still another can mould coöperative, flexible, purposeful, turn-taking, we-spirited persons. The group climate that produces such effects is created by the resultant of a number of group properties which can be combined in various ways, among which are the leadership style of the teacher or that of those who function most as group leaders, the degree of cohesiveness, which has already been mentioned, the group-member-skills, the suitability of the group process for the task in hand, the techniques employed by the teacher to satisfy his ego and other needs, and the tension-release patterns used by the group.[12,18]

10) The reasons for the occasional failure of project methods,

and other teaching procedures which depend upon effectively functioning groups often lie in the ineffective use of group problem-solving methods, or in the unskillful handling of group procedures. Groups can help themselves to mature and improve their ability as a learning or producing team by diagnosing their own failures and planning ways of repairing their own deficiencies. Students of group development have devoted much attention to methods of group diagnosis, ways of presenting the findings to a group, and methods for alleviating a group's procedural difficulties.[8]

11) Certain forms of classroom behavior may be recognized as mechanisms developed for relieving tensions somewhat similar to those employed by an individual in relieving his tensions. For example, they employ patterns of group behavior which help avoid difficult tasks or unpleasant situations. These mechanisms are often difficult to identify since they may either be wrongly perceived by the teacher as signs that the group is keeping busy, or they may be accepted as the usual troubles one gets into by the use of committee methods.[13]

12) Difficulties in the transfer of verbal learning to social behavior, can often be overcome by the use of that form of rôle-playing referred to as reality practice, in which the participants try-out the behavior they are expected to use in a situation from which all threat has been removed. Inhibition blindnesses, or fears of 'learning' certain content, or behaving in unaccustomed ways can be removed by the use of a 'cultural-island,' a situation where new group standards are generated while away from the source of the inhibitions. This procedure is effectively used in excursions, conferences, summer camps, and other group activities in which the person is under the pressure of group standards that are different from those at home, and so he dares to adopt forms of behavior which might be quite desirable for him, but which he might hesitate to try out in his accustomed environment for fear of adverse criticism.[7]

Thus we can safely accept the view that group phenomena definitely affect the progress of learning, as well as the kind of learning that takes place. The educational significance of this view derives from the fact that the pupil's attitudes as well as his behavior patterns are modifiable. Increased motivation in participating in the classroom activities, and consequently in learning, derives from several different potential sources in a group atmosphere where good mental hygiene prevails.

Three such potential sources of increased motivation will be considered. The first of these sources lies in method of *goal determination*—the extent to which the goals of the class are determined by the entire group including both pupils and teachers, in a truly co-participant sense. When this procedure is followed, the child will feel that he has some control over his own destiny and, therefore, is able to accept the group goals which he helped select as being his own personal goals. They are things which he himself wants to do and, therefore, he is more likely to follow through on them. The absence of such codetermined objectives does not mean the absence of group standards, but some of these standards are not likely to be the ones which the teacher would choose, or the ones which best promote learning. Such group standards as the 'gentlemen's mark' of C, and the group rejection of the student who is too 'eager,' are familiar to all. Thus group standards in a classroom may inhibit good learning as well as accelerate it.

The second source of increased motivation lies in the extent to which the teachers and the pupils build a *supportive atmosphere* in the classroom, one which helps each child to realize that he is *an accepted group member. When this condition maintains,* each child has his own 'area of freedom,' within which he is free to make his own decisions. This area can often be much wider than is ordinarily supposed by teachers who are constantly making pupils' decisions for them. Although the group may not approve of everything a pupil does, it still accepts him as a person. In this kind of an atmosphere the child is able to develop a greater feeling of security with his fellows. In addition—and this is the important contribution to learning—he is likely to feel freed from personal threat and criticism and, therefore, more willing to go ahead and try new things without fear, realizing that if he fails he will not be rejected either by the class or by the teacher. Thus failure can be a very positive learning experience because, once the emotional threat is removed, the child can look at his abilities and limitations far more objectively and with greater awareness of what next steps are required for his learning. It would seem that little learning can occur if the child is denied positive opportunities to make errors.

A third potential source of increased motivation lies in the extent to which the various members of the class are accepted as *participating members.* When they are so accepted, each can benefit from the knowledge, skills, and abilities of all the other members. They are no longer dependent primarily or solely on the teacher for all information and guidance. Besides offering the possibility of the development of broader understandings, this gives to each pupil the opportunity to be a contributor to the group, and the classroom becomes, then, a situation for mutual exchange, for mutual sharing. Research is beginning to show the increased productivity of groups which have this coöperative pattern of relationship.[3] Goal determination by the group, a supportive atmosphere, and a participating membership, then constitute three conditions of group organization of great effectiveness in developing motivation which contribute to the promotion of effective learning.

THE RÔLES OF A TEACHER

What can the teacher do to develop and maintain these conditions conducive to learning? There are three fundamental rôles which cover the things a teacher does. Actually these are not discrete parts of the teacher's job, but they do carry quite different implications. The rôles that will be discussed are the following: (1) the instructional rôle, (2) the rôle of the democratic strategist, and (3) the rôle of the therapist. Following this, we will ask how the teacher selects the proper rôle, and how the actual operation of this rôle can be evaluated.

First, the *instruction rôle.* It is obvious that the concept of what a teacher should do has changed over the years. To the Hoosier schoolmaster the matter was quite simple. He was the drill sergeant. The cadence of recitation was akin to the sound of marching feet. As master of the drill, he called the steps. This teacher also held the rôle of academic authority; not only did he choose the school experiences, but he was also revered for his great storehouse of information. His very person was the embodiment of learning, and he was categorically right. This fundamental instructional rôle has mellowed with the years. Now the teacher does not always have to know. He operates as an adult with superior learning to be sure, but serves more as a resource person explaining, telling, and demonstrating.

His drill-master's uniform has been exchanged for the Socratic garb, for his instruction is more concerned with fostering the students' power to think and reason. This major 'informational rôle' of the teacher is often discussed and is perhaps quite well understood. But it should be clear that this rôle itself is not exclusively the property of the teacher. At times, especially as the content of the course falls within the experience of the students, the class members share or take over the instructional rôle. As we come to understand more about the dynamics of the classroom, we realize that the way in which this rôle is handled by the teacher has important effects on the total learning situation.

A second major rôle which the teacher must play is that of democratic strategist. This has been discussed by other writers under the heading of "group formation." With the goal of pupil participation the teacher must provide the occasion for the introduction of processes to facilitate teacher-pupil planning. To play this rôle successfully two things are required: a high regard for democratic values, and their implications, and a high level of psychological insight into group factors and individual personality. In the rôle of a democratic strategist, the teacher helps the group utilize various methods of progress evaluation, *and the information about their progress which they secure.* He further helps them see and clarify their accomplishments, blocks, and failures, as well as the values in democratic group action. Thus the task is more than that of being merely an exponent of democratic education. This rôle becomes one of activating democratic processes by helping the class to experience democratic goals and relationships in the design of their everyday classroom experiences.

Understanding the dynamic forces which are affecting the class as a group and those which the techniques bring into play makes possible a contribution to democratic learning because our democratic ethics have established the educational goals and values. Techniques are selected in terms of their potentiality for contributing to the democratic goals of the group at the particular time. It should be pointed out that on the basis of a different set of ethics for the same conditions in a group, different techniques would be selected in order to achieve the goals determined by these differing ethics. However, since it is a contribution toward democratic learning that is desired, it is essential that teachers become as skilled as possible in understanding and working with their classroom groups. For a lack of such skill is likely to result in conditions which are quite the opposite of democratic, even though democratic techniques were supposedly being used. Democratic techniques do not exist *per se;* a technique is democratic only to the extent that it serves as a means to help the group achieve its democratic goals at a particular time. For example, the democratic technique of voting has been used as a very effective method of imposing some small minority opinion on the group.

A third important rôle of the teacher can be subsumed under the title of therapist—a combination of clinician and group worker. Lest someone remonstrate at this obligation, let him be reminded that, willingly or not, every teacher plays this rôle. Sometimes it is somewhat separate from other functions, but more often it is embedded in the classroom life while other functions predominate. No teacher avoids being a group worker, although some are more successful than others and some do crude jobs to be sure. The rôle of therapist implies group management to the end of helping all of the children toward individual and social adjustment. This means a degree of permissiveness, the establishment of rapport with each child, and the conduct of the work without the teacher's ego becoming involved. Such masterful, objective, 'impersonal' human relationships are hard to come by. No one person is able to meet the differential needs of thirty-five or more children and serve as a cushion to soften the blows of harsh reality dealt out by the child's peer culture. But one tries. To do this the teacher must so act as to be the implicit embodiment of an acceptable code of behavior. Time and time again the mores of mental hygiene are illustrated as the teacher relates to the children, to their feelings and to their problems.

It is through the supportive atmosphere previously discussed that the teachers' therapeutic work is carried on. In a conflict situation pupils may come to the teacher as a judge or decision maker. The case need not be handled arbitrarily, but it must be handled. Teachers can never be neutrals but are continually interpreting 'the law' as it applies in individual cases. In the therapist rôle, the teacher shares insights concerning human behavior, helps to get at causes of conflict and to find methods of resolving it. Sometimes the teacher serves this end by just being a friend, or he may provide, or himself be, an example with whom the child can identify in the Freudian sense. At any rate, the teacher must be an expert in human relations, understanding both the group and the individual.

In general teachers play this rôle least adequately of all. They tend toward being moralists, policemen, or punitive agents expecting good character to be developed by decree. While we have much to learn in applying the therapist rôle to the teacher, we already understand enough to know that such a playing of the rôle spells failure. The reason for such failure may often be that the teacher, having personal needs, tends to exploit the situation to satisfy these needs. We have in mind the need to be loved, the desire to avoid conflict, or pressures from latent hostility as examples. A very common attitude is the desire for dependency, where the teacher is happy if the students remain attached and dependent. Redl[14] has written a very interesting paper approaching this from a slightly different angle in which he shows how teachers tend to orient the whole atmosphere so that it plays into a masochistic or sadistic syndrome, to take only two examples. This is a complicated study in depth psychology, fraught with controversy. But it is not without point to us.

SITUATION AND CHOICE OF RÔLE

From the point of view we have been discussing, it will be seen that there is no single complex of rôles a teacher plays. The different legitimate objectives of a classroom demand different emphases. Certainly groups of children differ in their leadership qualities, and other individual and group factors need to be studied and understood. The question the teacher would then ask is: "What technique will contribute most effectively, in terms of the dynamics of my class at this time, to the goals and values which are held by the class (or myself, depending on who determines the goals)?" Two things are needed in selecting the techniques: (1) a knowledge of the dynamics of the technique itself, and (2) a knowledge of the goals and values of the group.

Knowledge about groups will help materially in gaining an understanding of the dynamics of a particular technique, and of the kinds of forces in the group which it brings into play in a positive (or negative) manner under specified conditions. To know these dynamics is important. Otherwise the teacher may fall into the trap of thinking that certain techniques are 'good' *per se,* forgetting that a technique will contribute to the group only as it is able to draw on the positive forces present in the group at the time. If the condition of the group is different at a particular time, the 'good' technique may bring out all that is 'bad' in the group, causing him to wonder why it didn't work, or to blame the group for 'not coöperating.'

SOME TYPICAL CUES FOR RÔLE CHANGES

How is it possible to determine which rôle to play at a particular time? What are the characteristics of a group which will serve as cues for shifting rôles? One such cue is group 'apathy.' If the group is lethargic and passive, one must start searching for reasons. Is it the course content? The teaching methods? A general atmosphere of repression? Children who do not become boisterous at times are living under the control of teachers who are misers of freedom.

Another cue is to be found in the rapidity of 'spread of disorder.' In a group with adequate morale and goal involvement, disturbances do not spread easily. If one child upsets the room, individual work with that child is, of course, indicated. But more important is the signal it gives about the group condition. If a 'bad actor' is a source of rapid contagion, the bond of common purpose must be weak indeed. This condition may be caused by such a simple thing as the need for a change of activity due to a requirement for overlong attention to a specific task. It may be a tension for muscle discharge, or it may go far beyond this to a fundamental dissatisfaction with the teacher behavior.

Other cues for further diagnosis and rôle modification include the presence of isolates, cliques, scapegoating, exclusiveness, extreme competitiveness, and the like. How much do teachers know about diagnosing these things? Indeed, how much help can educational psychologists give? Once the teacher really understands the situation and appreciates its deeper aspects, the rôle complex to meet the situation can usually be found. The task of the educational psychologist is to see that teachers are so trained that they will understand the dynamics of that situation.

Understanding more about the dynamics of groups helps the teacher in a variety of ways toward increasing his effectiveness in the rôles that are appropriate in different situations. As more is learned about the theory and research on groups, new ways of thinking about the classroom situation will at first be gained, ways which may have been overlooked before. The importance of effective communication will come to be recognized in giving instructions and in expressing ideas. The relationships between the various pupils in the class will be studied, how they feel about each other, and the leader-group relationships, and gradually the teacher will become aware of his own behavior in the class and the kinds of effect it has on the pupils.

Of course, it is not easy to take one's knowledge into the classroom and become immediately aware of these complex interrelationships. Often it takes considerable training in observation and experience to be able to see, especially at the time it is happening, what is occurring in the group and what its causal relationships and potential effects are. The transition from 'book learning' to 'observation skill' is a difficult one to make, but it must be made if knowledge about groups is to contribute to teaching effectiveness.

EVALUATION

"How does one know one has effectively employed the correct rôle? Were the results in the true psychological sense, those which were described? Was there progress by the individual or the class in the direction of the goals which had been established, and was this progress as great as it might have been if some other teacher rôle had been used, or if this present rôle had been carried through more effectively? And were the dynamics of group relationships improved as a result of this particular rôle? Is the class in the 'healthier' condition and more ready to take forward steps toward whatever new goals may be established, or have they achieved some of their important goals at great cost to themselves and to their interrelations in the group?

Information about these questions can come from different sources. The teacher, by employing the same sensitivity and observational skill used in individual diagnosis of pupil difficulties will become accustomed to diagnosing the group. An examination of the condition of the group will be an examination, at least in part, of the way in which the role previously employed affected the group and their response to it. Diagnosis and evaluation, then, go hand-in-hand as a continuing process for the teacher. Evaluation of a previous step, in a large measure, provides the cues for the next step, and for the choice of the role to be employed.

Of course, the teacher is not the only source of evaluation data in the classroom. To overlook the students' contributions is to disregard not only a most important source of information, but it is to deny the students the opportunity of evaluating their experience in the class which is the basis for making decisions to improve themselves as individuals and as class members for future work. It is not an easy task, obviously, to carry through an effective evaluation as a group, but the process may be a most valuable educational experience for all.[10]

A third source of evaluation data depends on the availability of outside persons who could be called into the classroom as observers. Someone who is not himself involved in the group is often able to note many important situations which the person who is trying to carry through an effective teaching job almost necessarily overlooks. The outside observer—whether he be a supervisor, principal, fellow teacher, or trained clinician—can note these situations. And to the extent that he has the personal skill in his relationships with those individuals to discuss his observations with the teacher freely and acceptingly, he can be of service in increasing the teacher's own skill in the classroom. He may also take the next step and open his insights to the group as a whole, helping them to see and comprehend more fully the processes of group interaction.

REFERENCES

1) William C. Bagley. *School Discipline.* New York: Macmillan, 1917.

2) Lester Coch and John R. P. French, Jr. "Overcoming Resistance to Change." *J. Human Relations*, Vol. I, 4, 1948.

3) Morton Deutsch. "An Experimental Study of the Effects of Coöperation and Competition upon Group Process." *J. Human Relations*, Vol. II, 3, 1949.

4) ———. "A Theory of Coöperation and Competition." *J. Human Relations*, Vol. II, 2, 1949.

5) Leon Festinger, Kurt Back, and Stanley Schachter. *Social Influence* (pre-publication book manuscript, mimeographed) Chapter VI, "The Operation of Group Standards" and Chapter X, "A Theory of Group Structure and Group Standards."

6) John R. P. French, Jr. "Organized and Unorganized Groups under Fear and Frustration." *Iowa Studies in Child Welfare, Studies in Vector and Topological Psychology*, Vol. 20, 1944.

7) Charles E. Hendry, Ronald Lippitt and Alvin Zander. *Reality Practice as Educational Method.* Psychodrama Monographs, No. 9, 1944, pp. 36. *Sociometry*, May 1944.

8) David Jenkins. "Feedback and Group Self-evaluation." *J. Social Issues*, Vol. IV, 2, 1948.

9) Kurt Lewin. "Group Decision and Social Change." In Newcomb and Hartley, *Readings in Social Psychology*, New York: Henry Holt, 1947, pp. 330–344.

10) ———. *Resolving Social Conflicts*, New York: Harper Bros., 1948, Chapter IV.

11) Ronald Lippitt. *Training in Community Relations*, New York: Harper Bros., 1949, pp. 286.

12) Ronald Lippitt and R. K. White. "An Experimental Study of Leadership and Group Life." In Newcomb and Hartley, *Readings in Social Psychology*. New York: Henry Holt, 1947, pp. 315–320.

13) T. Main and M. Nyswander. *Some Observations on the Third National Training Laboratory in Group Development.* (mimeographed)

14) Fritz Redl. "Group Emotion and Leadership." *Psychiatry*. Vol. V, 4, 1942, pp. 573–596.

15) Stanley Schachter. *Deviation, Rejection, and Communication.* Doctoral dissertation, University of Michigan.

16) John Thibaut. *The Relationship of Group Cohesiveness to Inter-Group Status Differences.* Doctoral Dissertation, Massachusetts Institute of Technology, 1949.

17) Alvin Zander. "Resistance to Change—Its Analysis and Prevention," *Advanced Management.* Vol. XV, 1, Jan. 1950, pp. 9–12.

18) ———. "Within the Bonds of Freedom." *Childhood Education*, Vol. 24, 1, Sept. 1947.

2) WHAT EDUCATIONAL MEASUREMENT IN THE EDUCATION OF TEACHERS?

WALTER W. COOK
University of Minnesota

In determining what educational measurement has to contribute to the education of teachers we shall attempt to answer three questions: (1) How can measurement improve the classroom learning situation? (2) What are some of the most important knowledges, skills, attitudes, and abilities which the teacher needs with reference to measurement? and (3) How can these needs be most effectively met in the teacher education program?

HOW MEASUREMENT AIDS IN IMPROVING THE CLASSROOM LEARNING SITUATION

The most important characteristic of a favorable learning situation is a strong ego-involved drive on the part of the learner to acquire the various socially approved behavior patterns with which the school is concerned. To assume that learning activities can be *self-sustaining* and ends in themselves is perhaps one of the most common errors in educational thinking. Such activities are in reality the means of acquiring an increasing realization of social status, prestige, and other basic ego satisfactions. What the learner really strives for is social position with security, prestige status with his peers, favorable attention, and recognition of his special virtues, abilities, and accomplishments. He also wants success in what he undertakes, some adventure, and a progressive broadening of significantly related and meaningful new experiences. A good school or a good learning situation is one in which these fundamental cravings of the individual are satisfied through educative experiences.

The eight-year-old who gets up at an early hour to practice reading a story he has volunteered to read to the class, the algebra student who burns the midnight oil with the really tough problems, the extra hours spent on a special report, the football squad at its gruelling practice, and the boy who secures a paper route to earn money to buy a trumpet, to take lessons for six months, to get to play in the band and wear a flashy uniform, are all examples of the social prestige factor in the learning situation. The problem then is to make learning highly satisfying in this sense. It should not be preparation for prestige getting—it should be prestige getting.

Certain elements of the classroom situation which are related to the prestige goal and which may be improved through the proper use of measurement are:

1) Classification and grouping. The first consideration is that students should be classified in such a way that no embarrassment or stigma is attached. The student should be among his recognized peers if status in the group is to be sought. This primary grouping must be largely in terms of overall physical and social development. The secondary grouping must be in terms of common achievements and learning needs. Group unity and cohesiveness are essential. These result from having common goals, common understandings, common efforts, common interests, common difficulties and common achievements.

2) The individual student should be placed in reading materials and problems comfortable for his level of ability. Success and the means of attaining it must be emphasized. An optimistic and encouraging atmosphere should prevail in the classroom.

3) Individual instruction with recognition of specific accomplishments and attention to specific deficiencies should be possible.

4) Objective measures of achievement and progress are effective stimulants to further efforts.

5) The student should have a clear conception of what he is attempting to accomplish. The objectives must be clear. He should also know what his specific errors and misunderstandings are. The means of self-appraisal and identification of needs should be provided. Shortcomings should always be considered as something to overcome, not ridiculed.

6) The teacher should have an economical method of determining the interests, problems, aptitudes, abilities, physical condition, social adjustment and self-adjustment of each student.

7) The goals of each child should tax his abilities and yet be reasonable for him. There should not be competition among students unequal in ability.

WHAT THE TEACHER NEEDS TO KNOW ABOUT MEASUREMENT

Instruments of educational measurement are simply the means by which the quantitative aspects of human behavior are observed with greater accuracy and economy of time. To the extent that such instruments are valid and conform to the principles of quantitative logic it becomes possible to know with greater exactness the relationship among the various aspects of educational procedure the aptitude of learners, and changes in human behavior. The purpose of this being to make possible more accurate prediction and control in the educational process.

To use measurement effectively in improving the learning situation the teacher needs to know not only how to apply the instruments and read the scores meaningfully, but also how to modify the learning situation in terms of what the scores reveal. The teacher must have a bowing acquaintance with the fund of research based on measurement and commonly designated as individual differences, child development, learning theory, mental hygiene, and group dynamics. Measurement is simply a tool to increase the accuracy and efficiency of observation; its value depends upon the insights of the user.

The following itemization of what a teacher needs to know in order to use measurement effectively is suggestive, but certainly not exhaustive. The teacher needs to know:

1) What standardized tests are most commonly used in the subjects and at the grade levels he proposes to teach. He should have taken the tests, administered the tests, scored, and interpreted them.

2) How to compute percentile ranks and standard scores and interpret them with reference to equal units on the base line of the normal frequency curve. How to convert a standard score to a percentile rank.

3) How to compute age scores, grade scores and intelligence quotients. The advantages and disadvantages of each type of score.

4) How to compute, interpret, and use properly the following statistics: median, mean, standard deviation, product moment coefficient of correlation and simple tests of significance. (The computation of these statistics from a frequency distribution is not necessary and tends to inhibit understanding. Computations should be kept simple with emphasis on understanding.)

5) How to compute and interpret coefficients of reliability and validity with understanding of the factors which influence each.

6) How to interpret a test score with reference to the standard error of measurement.

7) How to construct an objective test with attention to the following details: (a) stating objectives in terms of changes in behavior; (b) organizing subject matter through which these objectives are attained; (c) the advantages and disadvantages of the various types of objective items; (d) the proper sampling of objectives and subject matter; (e) the answer sheet; and (f) the scoring key.

8) How to compute by at least one simple method the per cent accuracy and discriminating power of test items; what the relationship of these two measures are to the form of the distribution of test scores; and their influence on reliability and validity.

9) How to construct individual and class educational profiles to show development over a period of years.

10) How to prepare for and conduct an interview with a parent, analyzing the achievement and behavior characteristics of a pupil. What one should learn from a parent in such an interview.

11) What to anticipate when a battery of tests is administered to a class. For example, that the four-standard-deviation-range of ability and achievement at the first-grade level is approximately four years, at the sixth-grade level it is eight years, and at the high-school level it is approximately ten years. The teacher should know a simple method of determining this for any grade level. He should also know that the typical pupil will show approximately eighty per cent as much variability in his various traits as the class does in the various abilities.

12) That age groups show no more variability than grade groups even when a considerable amount of retardation is practiced. For example, if the beginning seventh grade were to be made up of all the twelve-year olds in the school the variability of this group would not be significantly greater than the seventh grade which exists with its typical range of about five years in chronological age.

13) That a high rate of failure simply reduces the general intellectual and achievement levels of a school through the hoarding of slow learning pupils.

14) That where pupils are stimulated to maximum development in a given area the best predictor of growth in that area is

past achievement, not intelligence. For example, the best predictor of achievement in arithmetic is past achievement in arithmetic, not the intelligence test score.

15) That pupils should be grouped differently for instruction in each learning area. For example, not only will the grouping for reading be different from that for arithmetic, but the grouping for basic reading instruction will differ from the grouping for reading in connection with unit activities and that still different bases for grouping will prevail in the remedial reading and free reading programs.

16) That not all the important aspects of child development are now measurable and hence other means must be used to assess achievement in many areas.

17) That some learning is of a highly temporary nature while other learning is relatively permanent. The tests used should emphasize the permanent learnings.

18) That readiness tests, diagnostic tests, and to some extent achievement tests are based on a systematic approach to errors and learning difficulties, hence by studying these tests the teacher will develop his ability to observe significant behavior and become less dependent upon such tests.

19) That the use of measurement makes possible the child *development approach to education as opposed to the subject*-matter-to-be-covered approach.

20) That because of the variability of classroom groups, the grade levels at which certain knowledge, skills, and abilities should be learned cannot be determined with any degree of specificity.

21) That since life outside the school recognizes and rewards a great variety of aptitudes and combinations of aptitudes, the school should do the same.

22) How to apply at least one readability formula to reading materials.

23) That determining pupil status in a given developmental area and adjusting instruction to status is good teaching procedure.

24) That measurement as a motivating condition has three functions in the learning process: (1) The energizing function, to increase the general level of activity and effort. (2) The directive function, to direct the activity of the learner into desirable

channels. (3) The selective function, to determine the responses which will be fixated and the responses which will be eliminated.

25) That the major functions of a comprehensive testing program are:

(a) To direct curriculum emphasis by: (1) Focusing attention on as many of the important ultimate objectives of education as possible. (2) Clarification of educational objectives to teachers and pupils. (3) Determining elements of strength and weakness in the instructional program of the school. (4) Discovering inadequacies in curriculum content and organization.

(b) To provide for educational guidance of pupils by: (1) Providing a basis for predicting individual pupil achievements in each learning area. (2) Serving as a basis for the preliminary grouping of pupils in each learning area. (3) Discovering special aptitudes and disabilities. (4) Determining the difficulty of material a pupil can read with profit. (5) Determining the level of problem-solving ability in various areas.

(c) To stimulate the learning activities of pupils by: (1) Enabling pupils to think of their achievement in objective terms. (2) Giving pupils satisfaction for the progress they make, rather than for the relative level of achievement they attain. (3) Enabling pupils to compete with their past performance record. (4) Measuring achievement objectively in terms of accepted educational standards, rather than only by the subjective appraisal of teachers.

TEACHING EDUCATIONAL MEASUREMENT

To achieve the objectives outlined above will require no less than the equivalent of five quarter hours of college work, three hours per week of lectures and readings, paralleling four hours of laboratory work. The course should be one of the first if not the first in the professional curriculum because: (1) It is systematic, has definite content, and requires strenuous intellectual effort. (2) It provides readiness for professional study and the reading of educational literature. (3) It introduces the student to the school and individual pupils in a concrete meaningful way. (4) It avoids the stigma commonly attached to beginning courses in education resulting from their highly generalized laboring of the obvious and the inconsequential.

Considerable individualization of instruction in measurement is desirable. The varying capacities of students, fields of specialization, levels at which teaching is planned, interests and readiness of students require that there will be some differentiation in *readings and laboratory work.*

The students should be placed in contact with elementary- or high-school pupils as early as possible. Each teacher candidate should have practice in administering, scoring, interpreting, recording and reporting various measurements for at least one pupil. By bringing together these individual studies the problems of both the individual pupil and the entire class can become matters of study and discussion. The purposes of measurement and statistics in the work of a teacher are made obvious from the beginning through such procedures.

Students should have experiences in reading educational literature which involves quantitative and statistical interpretation. The professional sequence in teacher education should make demands upon the student for his knowledge of quantitative, objective, and measurement concepts in reporting observations in student-teaching, in reading assignments, and in understanding lectures.

The students should have experience in constructing various types of teacher-made tests. The administration of such tests *may of necessity be delayed until the period of student-teaching. Emphasis should be placed on the objectives to be measured and* the refinement of tests on the basis of data collected through their administration.

Much of the detailed knowledge necessary for measuring readiness or making a diagnosis of needs in reading, arithmetic, English, etc. will of necessity be taken up when the student studies method in these specific areas.

The nature of the course can be most concisely described by outlining in parallel columns the approximate sequence of topics in the lecture and laboratory periods.

Lectures and Readings 3 hours per week	*Laboratory work, Problems and Observations*, 4 hours per week
1) The nature and extent of individual differences in the school.	1) Observing the administration of group tests.
2) Trait differences, and the limitations of over-all homogeneous grouping.	2) Observing the administration of individual mental tests.
	3) The scoring of a test battery.

Lectures and Readings
3 hours per week

3) The effects of various policies of promotion and grouping.
4) Teaching objectives as related to temporary learning, permanent learning, and homogeneity of achievement.
5) Administrative procedures necessary for meeting the educational needs of individual children.
6) Curriculum policies necessary for meeting the educational needs of individual children.
7) Educational objectives as related to measurement procedure.
8) The logical aspects of educational measurement. What measurement is, how it differs from enumeration, evaluation, judgment and testing. Derived scores, validity, reliability, sampling and objectivity.
9) Percentile scores.
10) Standard scores.
11) Age scores.
12) Grade scores.
13) Intelligence quotient.
14) The meaning of a coefficient of correlation.
15) The meaning and measuring of test validity.
16) The meaning and measuring of test reliability.
17) Factors which influence test validity and reliability.
18) The procedure in constructing an objective test. Statement of objectives. Outline of subject matter. Advantages and disadvantages of various types of items. The answer sheet and scoring key.
19) Determining item difficulty.
20) Determining discriminating power of test items.
21) The proper selection of items

Laboratory work, Problems and
Observations, 4 hours per week

4) Making educational profiles of individual pupils and studying growth records.
5) Studying the educational needs of a pupil by observing samples of work, permanent records, and observation of behavior.
6) Observing classes in which pupils have been grouped in terms of learning needs. Observing materials used, problems, procedures, and objectives.
7) Computation of percentile scores. (Using actual test scores)
8) Computation of the mean.
9) Computation of the standard deviation.
10) Computation of standard scores.
11) Problems involving the conversion of a standard score to a percentile rank.
12) The normal probability curve.
13) Computation of the product moment coefficient of correlation.
14) Predicting achievement in reading from reading readiness scores. (This for primary teachers, other problems of prediction for other teachers.)
15) The standard error of estimate.
16) Tests of significance.
17) Constructing an objective test, stating objectives, outlining subject matter, building items, making an answer sheet and scoring key.
18) Administering the test when possible.
19) Computing item difficulties.
20) Computing an index of discrimination.
21) Revising the test in terms of item difficulty and discriminating power.

Lectures and Readings
3 hours per week

for a test with reference to difficulty and discriminating power.

22) Correcting scores for chance success.
23) The criteria of an ideal testing program.
    Achievement tests
    Prognostic tests
    Readiness tests
    Diagnostic tests

Laboratory work, Problems and
Observations, 4 hours per week

22) Observing readiness testing and follow-up procedures.
23) Observing diagnostic testing and remedial procedures.
24) Observing a good teacher-parent conference.
25) *Determining the* reading difficulty of a book.

3) HOW CAN THE PSYCHOLOGY OF DEVELOPMENT IN INFANCY AND CHILDHOOD HELP TEACHERS?

WILLIAM F. BRUCE

State University of New York, State Teachers College,
Oneonta, New York

The term, 'psychology of development,' is used here to emphasize the study of 'relations' among the facts and factors of development. Educational psychology becomes more scientific and sound as it exposes relations between various aspects of development, such as those between physical growth and social acceptability, between emotional attitudes and intelligent response in test situations, or between social class status and motivation in school. The thousands of factual items discovered by recent investigations in such fields as physical growth, intellectual attainment, and sociological studies of youth in rural and urban communities, can so fill the time of instructors and students in professional courses that the psychological relations involved scarcely can be seen or even vaguely understood. If the psychology of development at any age level is to be helpful to teachers, it must reach far beyond the memorization of facts into actual comprehension of the processes of development.

One way of insuring that the teacher understands the psychological relations involved in development is to study growth with an eye upon the characteristic life-styles of particular persons. The life-style of a child is his way of walking, talking, meeting other children and adults; in short, his way of behaving as a whole person. In studying him we may look analytically at his physique, his health history, his family pattern, and his interests in and out of school; but our understanding of him and of the psychology of personality comes through relating these various parts to the individual's whole life-style. The process by which the teacher comes to understand the significance for education of the facts in a variety of fields is by alternating attention from the problem presented by the person's whole behavior to one or more parts of the situation and then back again with suggestions about how to guide the child. Thus the student comes to see the particular item as a member of a dynamic pattern of child behavior.

Since the personality of even a young child is exceedingly complex, the student of educational psychology will find his scientific sources classified under such titles as emotional energy and conflict, motivation from drive to purpose, and physical development. The student-teacher and his college instructor must not allow these analytic categories to break in pieces his concern for the well-rounded development of whole persons. These materials organized around particular aspects are to be used in discovering and understanding relations that may open the way for modification of the child's life-style through guidance.

Another way to keep the enlightening relations in the foreground of the teacher's study of the psychology of development is to think of the entire life-cycle rather than treating separately such periods as infancy, early childhood, childhood, adolescence, maturity, and old age. While the ten-year-old's most adequate preparation for maturity is living most fully today on his own terms according to his own present capacities, the teacher can scarcely understand him psychologically without seeing the relations between his behavior in earlier years and his present style of life. Nor can the teacher wisely guide the ten-year-old without a thoughtful look ahead toward the adolescence which is fast approaching and toward the maturity that lies beyond. Although the elementary-school teacher takes particular responsibility for children of a certain age level, he will understand the psychology of development at that age more fully as he sees it in the long context of an ideal life-cycle stretching through four score years or more.

A further warning emerges from the discussion of age-level periods, such as childhood and adolescence. Too often elementary-school teachers and secondary-school teachers in both pre-service and in-service training are faced with courses entitled for the former, "Psychology of Child Development," and for the latter, "Psychology of Adolescence." This legitimate difference in emphasis is apt to lead into drawing sharp lines of contrast as though the jump from childhood into adolescence came on a particular day and radically changed the style of life. Actually, although the first menstruation may be taken as a physiological criterion of the girl's leaving childhood and entering adolescence, her change in bodily form and rate of physical growth often becomes evident many months earlier, while her interest in boys may have likewise changed earlier or may be much delayed. The chief psychological principles helpful to teachers have to do with the continuity of life-style in the midst of growing up and with the great range of individual differences in rate of physical and social development. Since many elementary-school youngsters are physically and socially adolescents, and many high-schoolers are still childish in body or spirit or both, and because of the continuity throughout the life-cycle, all teachers may well study the psychology of development over a span of life inclusive at least of childhood and adolescence.

Since it is our responsibility in this discussion to emphasize childhood rather than adolescence,[1] we turn back to the beginnings of life to find in these early stages psychological principles helpful to the teacher. How much time and emphasis shall be placed upon considering the very beginnings of life involving the processes of physiological heredity and pre-natal growth? Here we run the danger, when we use a chronological-development approach, of getting halfway through the course before the child is born as we get immersed in the interesting facts of pre-natal development. The teacher does need, however, to relate the facts of the complexity of human inheritance and the unpredictability in particular cases of general capacity and temperament to the range of individual differences in the school class. Likewise, the pre-natal stage throws light upon the relations between endocrine-gland action and physiological growth and thereby later intellectual capacity. Birth injuries and partial suffocation may be mentioned as added evidence that the diversities in general capacity may have causal factors other than genetic differences. By linking the pre-natal and post-natal months the dual principle of differentiation-integration in physical growth and motor development may become clear enough on these relatively simple terms that it may be understood as a psychological process underlying much of our modern view of intellectual and social learning in later years. Thus we turn back to the very start of life because there the teacher of any level can reach certain understandings of development most effectively.

One problem that students of development raise insistently is the relative effect upon the whole life-cycle of the infancy period. Here the teacher needs much help from the educational psychologist in reaching a balanced view between extremes. Although the public generally and educators particularly have largely escaped from the clutches of the fatalistic hereditarians, they are plagued by other doctrines to the effect that the life-style is fixed beyond repair or modification in infancy. The behaviorists and sociologists have contended that conditioning factors and the cultural patterns early make the person what he is, and there he stays. Apparently, the teacher is offered no hope that the child can be guided nor that the culture can be improved. The Freudians have depicted how the oral, anal, and genital experiences of infancy along with much mistraining in our culture about toileting and sex have established a fixed personality which feels and sees life in unchanged ways even through the adult years. Other clinical studies of infants under impersonal institutional care have brought recognition of the need of cuddling the baby, if we are to have either healthful physical and social development or effective intelligence. Here again the class in educational psychology must select carefully and balance evidence judiciously if it is to get on with the study of children supported by a degree of confidence that some may mature adequately. On the other hand, *to refuse to face the evidence of the profound and far-reaching effects of infant experience submitted by the clinician,* the sociologist, and the psychologist of learning is to blind the teacher to the important relations between infancy and later childhood.

---

[1] See discussion of adolescence by Dr. Glenn M. Blair following.

Can the teacher of eight-year-olds get psychological understanding from studying the behavior of two's and three's in nursery school and four's and five's in kindergarten? The specialists in 'early childhood education' reply that this period is just the stage in which real understanding can come as the student sees these young children forming their individual life-styles in groups. Certainly, the numerous studies of the social development of young children throw a great flood of encouraging light upon the possibilities of development with skillful guidance in small groups. Much more of the favorable change in elementary-school practice during recent decades can be credited to leadership from below in the kindergarten than from above in the high school and the general college. The relation to the learning process of high energy output of run-about children, when released in activity among appropriate things for building, climbing, and picturing, and among other pre-schoolers and teachers with whom one can talk and play, eat and work, has been revealing. The teacher of older children will wish to move on, however, after careful attention to selected studies of these younger children which are most significantly related to the development of older children. The psychology of the young child, in spite of its vividness, should not too much delay the student whose responsibility lies with a higher age level.

What contributions to the teacher's work come directly out of the study of the psychology of development in childhood, say at ages six to twelve approximately? Space permits the mention of only a few fruitful areas.

The relations of the physical organism to the cultural environment is a continuing theme of psychological study at every age level. In this back-and-forth process the child becomes a person to respect as he takes on the language and customs of his home, of his peer group of age-mates, and of the wider community. By this process, which can be likened to a conversation between two people, sometimes harsh and sometimes friendly, the helpless, hungry infant becomes a child with a measure of intelligence and of social accomplishment. Fortunately, there are two sides to the process, so that he early begins to remake his family, his play group, his school class, and even his teacher. This process is often labeled 'interaction,' but it is more than friction between two cogwheels, more than the wear between auto tire and pavement. It is actually a 'transaction' in which a change occurs in both the individual and the society. This process of acculturation is creative, although some sociological discussions of cultural pressures and class status sound as if the child was merely being 'molded' by his culture. How the culture is transmitted to youth and at the same time is gradually transformed, especially in a modern democracy, is a central problem of the psychology of development.

The teacher today finds at hand many new techniques for studying the relations of the individual organism to the cultural environment. Sociometric techniques help the teacher to understand how the children feel about each other; who is accepted and who is neglected or rejected. The relations to learning situations can be traced, and teachers begin to build a vital psychology of their own. The studies of class status in typical communities give the alert teacher hints concerning the origins of learning failures and successes in family culture and language habits and in emotional feelings of inferiority, superiority, security, or mobility strain. The study of groups in action on the adult and adolescent levels is indicating some of the ways in which competition among the pupils in a class can be replaced by team work of the children, which is at once more rational and more democratic. Thus the modern teacher gains help from a deeper understanding of the child as a member of a class group, and teachers may also receive suggestions concerning the development of a class of individuals into a group that works dynamically with the teacher.

Some instructors make motivation the central theme of their teaching. Here the studies may range from the measurement of various biological drives in rats to what part of a child's day is given to the movies, radio, and comics in 1950, and on to the ethical purposes that begin to take form in the relatively mature twelve-year-old. In all this study of motives the important, but often neglected point, is the relation between the facts and the personality developing in individual children. Many pages have been written on the psychology of interests involving the collection of much data, in which the psychology of development ran much thinner than in that old study of young John Dewey on *Interest and Effort.* The relating of concepts such as interest and effort, freedom and control, external and internal discipline, drive and purpose, is more useful to the teacher than bookfuls of raw facts. The psychology of the development of motivation is still a fertile but only partially cultivated field.

After venturing into cultural inquiries that emphasize group processes and considering the possibilities of modifying interests, is it helpful to teachers to look at the individual again in terms of a personality or 'self' seeking maturity? Such an analysis of the 'self' into several more or less related phases aids the student in understanding children and incidentally his own more-or-less mature 'self.' Certainly, psychological distinctions arise out of considering the 'self' (1) which he has, (2) which he thinks he has, (3) which others think he has, and (4) which he thinks he ought to have. Using the contributions of philosophers, psychoanalysists, and lay observers the teacher may learn to distinguish among (1) the actual behavior of a child, (2) that child's own description of his behavior with its rationalizations, excuses, and escapes, (3) the biased evaluations of others, particularly the teacher's own estimate, and (4) the ideal or super-ego to which the child aspires. And before the teacher leaves the study of the psychology of development he may become concerned to understand himself more fully and to overcome some of the personal barriers that stand between him and accurate study and wise guidance of the children. Among these barriers within the personality may be mentioned a woodenness or lack of inspiration, extroverted showmanship, compulsive rigidity, persisting fatigue, continual fear for one's security, dependence upon others, and lack of self-discipline or organization in one's life. For the teacher, and all of us, are only part way along the road to maturity. We are still seeking an integration in which our four selves come into more intimate relation and closer unity with each other. As aids in this creation of mature integrity in the person we have on the one side the resources of child-analysis and self-analysis and on the other the ability to clarify the goals for personal development that are possible and desirable in a culture aiming to be more free, more responsible, more democratic.

If there have crept into these suggestions on the 'content' of the psychology of development certain assumptions about 'methods' of instruction in this area, we can only plead that content and method are as inseparable as organism and oxygen in breathing, as organism and culture in human development. Although considerable contact with children by the pre-service student is here considered essential, the relation between experiences with children and the organization of courses in the psychology of development presents many problems. How useful are case studies that are read, that are seen in movies, that are made through a period of study of actual children? How early shall child contacts, on the one hand, and logical organization of psychological subject matter, on the other, be introduced? How fully aware should the student become of the effects of group work upon their own personal development, their rapport with children, and their learning in this area of human relations? Such questions indicate the difficulties in the adequate education of teachers at present with the limitations of resources, personnel, and time. It does appear economical to aim at the teaching of concepts and generalizations, which are understood well enough to be transferred readily to new situations teachers will meet, remembering that the adequate understanding of these principles and policies can come only through experiences that have concreteness for the student. These concepts comprehend sets of vital relations which give them psychological power.

The mention of certain crucial problems in these pages is intended only as suggestive for other teachers of educational psychology. The selection of both content and method depends upon the class membership and institutional conditions. For example, in many colleges the student will come to the course in child development or educational psychology well informed on some of the biological, psychological, and sociological problems in our list. Each instructor needs to map out his own area of responsibility in relation to other fields. He must establish and modify from year to year his own pattern of instruction. Communication throughout the field of educational psychology is to be encouraged to the end that each instructor will have at hand a wide variety of suggestions from his colleagues across the country as well as within his own faculty. Then it becomes the responsibility of each instructor to develop the kinds of courses in education which will promote in the students whom he meets such an understanding of the psychology of development as will further the maturity appropriate for living in a democratic community.

4) WHAT TEACHERS SHOULD KNOW ABOUT THE PSYCHOLOGY OF ADOLESCENCE

GLENN M. BLAIR
Professor of Educational Psychology
University of Illinois

The question has sometimes been raised as to whether there is any unique content to the field known as the 'psychology of adolescence.' It has been argued that the adolescent is a human being and is subject to the general laws of human behavior the same as anyone else; and that one who thoroughly understands the basic general principles of development, learning, and adjustment will be amply qualified to deal with any adolescent problem that might arise. Those taking this viewpoint also might assert that all specialized courses in psychology overlap with every other course. The instructor teaching the course in general psychology will refer to the people about whom he talks as 'subjects'; the teacher of the educational psychology course will utilize the word 'pupil' in his discussions; the course in child psychology will employ the term 'child'; and the course in the psychology of adolescence will substitute the word 'adolescent' to take the place of such words as 'subject,' 'pupil,' and 'child.' In all other respects the courses I have just referred to would be very much alike if not indistinguishable. There is, of course, an element of truth in this type of argument. Courses in psychology inevitably overlap and reinforce each other in terms of basic theory. And, as often taught, the amount of overlapping may be so great that the distinctive contributions of a given specialized course may not be realized.

The present writer takes the position that what we know about human behavior is applicable to understanding the adolescent, but that the period of adolescence has special problems which require additional study. There is a growing body of valid information concerning the period of adolescence which is usually given but slight attention in general books and courses in psychology. Much of this is of the greatest importance to teachers who work at the secondary-school level. The purpose of this paper is to briefly call attention to some of this material. The special problems and principles of adolescent development to be discussed will be presented under the following four headings: (1) Adolescence—a transition period, (2) Adolescence and the culture, (3) Biological changes and the adolescent, and (4) Developmental tasks of adolescence.

ADOLESCENCE—A TRANSITION PERIOD

The teacher who works with individuals of adolescent age should realize that this period represents one of the chief transition stages of one's development. It is a time when the individual is attempting to make the switch from a childhood rôle to one of adulthood. This task of changing from a child to an adult is entirely new to the individual, and one which he will not have to face again during his life span.

During adolescence the individual is neither a child nor an adult. He is often treated in an ambiguous manner by his parents, teachers, and other members in his society. He is told one moment by his parents, for example, that he is too young to drive the family car; the next moment he is informed that he is too old to be carrying on certain childish antics such as crying. Roger Barker has stated that "for the adolescent, the overlapping situations of childhood and adulthood are paramount. Adolescents, by virtue of their frequently ambiguous and rapidly changing physiques, are often placed in marginal situations with the physiological, social and psychological behavior determiners of both childhood and adolescence acting upon them simultaneously."[1] It is very probable that the vacillating and unstable behavior often attributed to adolescents is due to the fact that they find themselves in new and overlapping situations where their rôle is not structured. Teachers need fundamental insight into the psychological processes involved in such situations in order to understand the adolescent and to help him define his own rôles more clearly. The topological theories of Kurt Lewin and his followers are particularly helpful in providing an understanding of the problems faced by adolescents during their transition from childhood to adulthood.[2]

---

[1] Roger Barker, et al., *Adjustment to Physical Handicap and Illness: A Survey of the Social Psychology of Physique and Disability*, Social Science Research Council Bulletin 55, New York, 1946, p. 33.

[2] See Roger Barker, et al., ibid, pp. 22–44.

ADOLESCENCE AND THE CULTURE

All human behavior is, of course, to a great extent influenced by the culture in which it is surrounded. Particularly marked, however, are the effects of culture upon adolescent behavior. The length of the period of adolescence, for example, depends upon the way society treats the adolescent. In some primitive societies children are transformed into adults almost over night. At thirteen or fourteen years of age they are given adult responsibilities and quickly weaned from the family. In our own culture in pioneer days it was not at all unusual for young people to marry at fourteen or fifteen years of age and take on family responsibilities. Such individuals probably had at the most one or two years to be adolescents. More recently, particularly during the depression years, it was not at all uncommon for young people to spend as much as ten years in making the transition from childhood to adulthood. During the recent war, great numbers of boys were automatically matured into adults upon reaching eighteen. At this age they left home and entered the service of their country, thus cutting short the usual period of dependence upon their families by several years.

The type of culture which surrounds the adolescent not only determines the length of the period, but produces well-known and distinguishable effects upon his behavior and personality. Margaret Mead several years ago documented the fact that young people in Samoa did not act like individuals of comparable ages in the United States. Because of the way they were reared these island boys and girls seemed to be lacking the symptoms of 'storm and stress' which are reputedly exhibited by our own youth.

Anthropologists and sociologists, who have studied the behavior of adolescents reared in different cultural settings in our own country, report great differences in the behavior patterns found. Allison Davis states that "lower-class culture, white or Negro, organizes adolescent behavior with regard to aggression, sexual relations, age rôles, and family rôles . . . into patterns which differ radically from those of middle-class adolescents."[1] His evidence shows that what is rewarding to a middle-class adolescent is not at all so to an adolescent reared in the lower-classes. Their attitudes toward schoolwork, what they fear, crave or cherish are distinctly different. Similar findings and conclusions are reported by Hollingshead in his study entitled *Elmtown's Youth.*[1] The material to be found in such studies is of the greatest significance to teachers who would understand the behavior of adolescents enrolled in their classes. Of value to teachers, also, are studies which have been made of the adolescent peer culture within a given social class. Such studies show how value patterns change with increasing maturity and how status within the peer group is attained. A good example of this type of investigation is Caroline Tryon's *Evaluations of Adolescent Personality by Adolescents.*[2]

---

[1] Allison Davis, "Socialization and Adolescent Personality," in 43rd Yearbook of the National Society for the Study of Education, Part 1, *Adolescence*, 1944, p. 209.

BIOLOGICAL CHANGES AND THE ADOLESCENT

An individual, of course, only has to face the facts of puberty once in his life. The adolescent's changed body creates new psychological problems which he must meet. There are the problems connected with voice change, adolescent acne, somatic variations, early and late development, etc. Especially important data regarding psychological problems resulting from physical changes and development during adolescence have come to us from the California Growth Study. Stolz and Stolz[3] have cataloged the physical disturbances and body variations which worry adolescent boys and girls, and Harold E. Jones[4] has reported on the psychological effects accompanying early and late development. His data show clearly that the early maturing girl is faced with more serious adjustment problems than the late maturing girl or the early or late maturing boy. The studies of Stone and Barker[5] on the attitudes and interests of premenarcheal and postmenarcheal girls have also given us increased insight into the effects of puberty upon the outlook of the developing child. The data show that among thirteen-year-old girls, those who have menstruated are far more mature in their responses than those who have not. Specifically, the postmenarcheal girls indicate more heterosexual interests, more interest in adornment and display of their persons, and less interest in participation in games and activities requiring vigorous activity than their premenarcheal age mates.

---

[1] A. B. Hollingshead, *Elmtown's Youth*, New York, John Wiley and Sons, 1949.

[2] Caroline M. Tryon, *Evaluations of Adolescent Personality by Adolescents*, Monographs of the Society for Research in Child Development, Vol. IV, No. 4, Washington: National Research Council, 1939.

[3] Herbert R. Stolz and Lois Meek Stolz, "Adolescent Problems Related to Somatic Variations," in 43rd yearbook of the National Society for the Study of Education, Part I, *Adolescence*, 1944, pp. 80-99.

[4] Harold E. Jones, "The Cycle of Puberty." Paper read at the Fifty-seventh Annual Meeting of the American Psychological Association, Denver, Colorado, September, 1949.

[5] C. P. Stone and R. G. Barker, "The Attitudes and Interests of Pre-

The teacher who would wisely counsel adolescent boys and girls should be in possession of these and many other facts which have to do with adjustment problems stemming from biological and physical change during adolescence.

DEVELOPMENTAL TASKS OF ADOLESCENCE

Individuals of every age group have problems which must be successfully solved if normal adjustment is to be maintained. The nature of these problems varies considerably at different periods of the life span. The infant or small child must master the complexities of learning to walk, learning to talk, and controlling the elimination of waste products of the body. The young adult has such special problems as learning to live with a spouse, rearing children, and securing a place for himself or herself in the occupational world. The middle-aged adult has such new problems as living harmoniously with adolescent children in the family, adjusting to physiological changes occurring during the menopause, if a woman, or adjusting to decline in general physical capacity, if a man. Still older people often have such unique problems to deal with as adjusting to retirement, or living alone after the death of a husband or wife. Robert J. Havighurst[1] has made a special study of the typical and somewhat unique problems facing human beings at various stages of their existence. He calls these problems or special learning activities 'developmental tasks.' He has suggested nine basic problems or tasks[2] which the adolescent must master if he is to achieve a successful adulthood. They are as follows:

---

menarcheal and Postmenarcheal Girls," *Journal of Genetic Psychology*, Vol. 54, 1939, pp. 27–71.

[1] Robert J. Havighurst, *Developmental Tasks and Education*, Chicago, The University of Chicago Press, 1948.

[2] *Ibid*, pp. 30–63.

1) Accepting one's physique and accepting a masculine or feminine rôle.

2) Developing new relations with age-mates of both sexes.

3) Attaining emotional independence of parents and other adults.

4) *Achieving assurance of economic independence.*

5) Selecting and preparing for an occupation.

6) Developing intellectual skills and concepts necessary for civic competence.

7) Desiring and achieving socially responsible behavior.

8) Preparing for marriage and family life.

9) Building values in harmony with an adequate scientific world-picture.

Problems, such as those just listed, are not entirely unique to the adolescent period, although some of them are. They are all, *however, tasks upon which adolescents must work during the* transition period from childhood to adult status. Teachers should understand in great detail the nature of these problems if they are to minister to the needs of youth, and succeed in reducing the stress and strain so common in adolescent development.

CONCLUSION

No attempt has been made in this paper to outline what teachers should know about psychology in general or even to catalog detailed information the teacher should possess regarding adolescent development. The purpose has been to suggest several areas of investigation concerning adolescents which receive scant attention in general psychological discussions of behavior and which have *especial significance for teachers who work with* individuals of adolescent age. It has been shown that the adolescent has special problems which he did not have as a child and which are somewhat different from those he will encounter as an adult. He has certain pressing needs which must be met, and a series of developmental tasks which he must master if he is to become a self-sufficient member of society. The successful teacher should understand what these problems are and how to assist the adolescent in resolving them. Knowledge of basic behavior theory plus specialized information regarding the period *of adolescence and the individual adolescent* are prerequisites for this task.

5) WHAT THE PSYCHOLOGY OF LEARNING HAS TO CONTRIBUTE TO THE EDUCATION OF THE TEACHER

G. LESTER ANDERSON

Dean of Teacher Education, New York City Colleges

It is the objective of this paper to explore the place of the psychology of learning in the education of the teacher. This objective is a part of a larger objective—that of determining a curriculum for teacher education. Our limited objective should be approached in terms of what it is,—a curriculum problem. The fairly well standardized procedures for curriculum work should be followed in the exploration of the place of the psychology of learning in the teacher education curriculum. To follow these procedures, answers would be found to questions of the following type:

1) What is the work which teachers do or should do?

2) What education (or training) does the teacher need in order to begin to do his work?

3) What are the experiences the teacher-in-training must have if he is to become an educated teacher?

When these questions are answered, we can then ask: Are any of the experiences which teachers-in-training should have subsumed, under present academic labeling, in a course or section of a course which can be called the psychology of learning?

Let us very briefly review what all of us might say in answer to these questions, as these answers would have relevancy in determining the place of the psychology of learning in the teacher's education. Because we can assume that learning has its most important contribution to make to the teacher's rôle *qua* teacher, this brief discussion will be restricted to what are essentially instructional and curricular aspects of the teacher's work.

The task of the teacher *qua* teacher can be briefly stated as follows: (1) To determine the goals of education and the contributions a given teaching area makes to these goals. This is the problem of objectives. (2) To determine the capacities of those being taught to attain these objectives. This is the problem of determining objectives for each individual learner. (3) To determine the experience which the learner is to have in order that he can attain his goals through learning. This is the problem *of setting the curriculum.* (4) *To organize these experiences* in terms of their scope and sequence so that learning will be at a maximum. This is the problem of determining the course of study. (5) To guide the learner in these experiences, again so that learning will be at a maximum. This is the problem of methodology. (6) To test the learner to see that individual and hence the educational goals have been attained. This is the problem of evaluation.

If the teacher is to perform these tasks, experiences which utilize a body of knowledge which we call the psychology of learning would certainly seem to be implied. Is it not axiomatic that a teacher must understand the learning process and be a guide or director of learning to the individual students under his tutelage. If this statement is acceptable we can then proceed to other issues. How teachers are to come to understand the nature of the learning process, the necessary depth of this understanding, and how they are to learn to direct learning wisely are the moot points. Let us address ourselves to some of the problems that thus emerge.

Psychology of learning should probably be part of a basic course in educational psychology. This course will be included in the teacher education curriculum at some point before work *in method or in practice-teaching.* We should like to make these assumptions about the teaching of the basic course in educational psychology.

1) General psychology should be a prerequisite.

2) Educational Psychology will maintain an identity as a body of knowledge in the teacher education curriculum.

3) Educational psychology will not be taught as a lecture-textbook course but will be taught in harmony with the best concepts of college teaching. Laboratory and field work, with case studies of pupils, observations of learning situations, etc., will be an integral part of the course.

4) The basic material in learning as a part of the course in educational psychology should probably be eclectic in character. Differences in theory have importance and significance for educational practice, but a detailed exposition of the different theoretical positions probably has little place in an undergraduate course in learning.

What shall be the content of the section dealing with learning in the basic educational psychology course? The following major divisions are suggested:

1) The general nature of learning.
2) The motivation of learning.
3) Transfer, retention, and forgetting.
4) Nature and direction of learning primarily motor in character.
5) Nature and direction of learning primarily intellectual in character.
6) Nature and direction of learning primarily social and attitudinal in character.

From the study of these topics the teacher-to-be should develop an understanding of basic concepts, and certain generalizations or principles about human learning, and come to see the relevancy of these for the performance of his responsibilities when he becomes a teacher.

Blair, among others, has shown that we are not yet ready to produce a definitive list of concepts and principles that can be generally agreed upon by educational psychologists which all teachers should understand. But we must begin systematically to produce such lists which can become basic elements of this section of the teacher education curriculum.

We have the following in mind as examples of learning concepts: motivation, incentive, need, interest, reward, practice, transfer, generalization, integration, inhibition, goal, effect, and experience.

We suggest the following as illustrative of principles which might be established:

1) Practice is a necessary but not a sufficient condition for learning.
2) An individual learns the responses which bring him to the incentives which satisfy motives.
3) Forgetting occurs not as a function of disuse of a particular behavior but because of the interference of new learned behavior with previously learned behavior.
4) Motivation without an appropriate incentive is ineffective to produce learning.

This approach, suggesting that teachers must understand concepts and principles which are definitely psychological in character, may seem formidable. Some would make the approach purely on a problem and direct experiential basis; that is, teachers would learn their psychology of learning by observing and working with children in learning situations, and would 'take from' the discipline which is the psychology of learning only those things which bear upon a problem the teacher-as-student faces in his work with children. This latter approach seems unsatisfactory if the teacher is to be trained as a professional practitioner rather than as a technician. It is only as a teacher masters the disciplines which bear on his work, as, for example, a physician masters anatomy, that he can be considered to have a professional education.

We should like to state immediately however, that instruction in this area should provide the teacher-in-training abundant opportunity, through direct experience with children as they learn, to see the concreteness and reality of the concepts and principles of learning which are abstractions, *per se.*

By way of conclusion, we should like to repeat how the psychology of learning makes its contribution to the valid performance of the teacher's task.

The psychology of learning should give some clues as to what is attainable on the part of the learner. Consequently, the psychology of learning can give reality to educational objectives. It can prevent educators from setting as goals the unattainable. This has too frequently occurred in educational history and has represented great wastage in educational endeavor, and frustration for those presumably being educated.

The psychology of learning plays a rôle in curriculum building in that it gives clues to the most appropriate organization of content and activity for purposes of learning.

Evaluation should be made in terms of behaviors to be learned. Again, through the psychology of learning the teacher may come to better understanding of what is to be evaluated and may better determine the appropriate techniques for doing this.

But the psychology of learning should play its central rôle in methodology. The problem of method is the problem of directing or guiding the learner in the experiences which will produce desired changes in his behavior.

6) THE STUDY OF INDIVIDUAL DIFFERENCES IN THE EDUCATION OF TEACHERS

FRANK S. FREEMAN
Cornell University

I

Every teacher knows, and so does every student preparing to teach, that people differ among themselves in respect to every human trait: physical, intellectual, emotional, social. In fact, educational philosophers and practitioners, beginning at the latest with Plato and Aristotle, have been aware of wide individual variations in mental ability and in the nonintellective traits of personality. Until relatively recent years, however, most educational philosophers and practitioners did little about the matter except, for the most part, to take a pre-deterministic and fatalistic attitude toward individuals and to eliminate those who appeared unable to learn a prescribed minimum in rigidly conceived curricula and courses of study.

It is less than fifty years that individual differences have been widely studied scientifically and subjected to quantitative measurement. In saying this, the pioneer work of Francis Galton nearly a century ago is not being overlooked; nor are the early exploratory investigations of Alfred Binet in the 1880's and 1890's, and the contributions of a group of American psychologists, notably James McKeen Cattell, prior to 1900. But extensive and intensive scientific psychological study of individual differences in man really dates from Alfred Binet's first intelligence scale of 1905.

After the publication of the first Stanford-Binet scale in 1916, and more especially after the program of psychological testing in World War I, we saw in this country an enormous number of studies contributing to the body of facts concerning human variations in respect to a great variety of psychological functions and educational achievement. The data of these studies were analyzed for the purpose of determining the causal factors of variations: differences due to sex membership; to age during the periods of growth and decline; to racial membership; to national origins; to immediate and to remote ancestry. Studies were made of variations due to environmental factors such as socio-economic status, childhood diseases, the amount and quality of schooling, cultural resources of the community, and the like.

Throughout these studies, either implicit or explicit, was the ubiquitous basic and primary problem of 'heredity and environment'; or what Galton so aptly named, the 'nature and nurture' problem. In a democratic society that has a system of universal public education, from the first grade through the university in nearly all states, and in a society that professes to provide equality of opportunity to all, and to foster all human capacities and human resources, the nature-nurture controversy is of fundamental importance not only to psychologists but to educators as well. Indeed, it is of fundamental importance to all who are concerned with education and with the society served by our educational institutions. For upon the resolution of that controversy, and upon the psychological and educational principles emerging from it, will depend to a very considerable extent our educational and social implementation of democratic educational doctrine; and, indeed, upon these emerging principles and practices will depend, to a significant degree, the mental health and personality development of many hundreds of thousands of American children and youths.

The nature-nurture controversy was considered important enough by the National Society for the Study of Education to warrant their devoting two Yearbooks to the subject; first in 1928 and again in 1940. In the first of these, the division was quite sharp as between the two opposing 'schools' of thought: the rather strict hereditarians and the environmentalists. Throughout the 1930's the number of researches on the problem increased and the debate continued—sometimes becoming more visceral than scientific and intellectual. Nevertheless, the gap between the opposing interpretations and developmental principles of the two groups has narrowed considerably, so that now nearly everyone attributes some significance to psychological nurture, and nearly everyone recognizes that genes are among the basic factors to be considered in mental development.

What has been said thus far is well known to all psychologists and educators who have studied the subject of individual differences. They know the field in respect to factual data, theory, and implications for social theory and educational practice. It is essential, also, that teachers, school counselors, and school administrators should be familiar with the background of this subject, with representative and significant studies in the field; and especially with the development and present status of theories regarding the nature and causes of individual differences: that is, differences among individuals and differences within each individual himself.

It is essential that teachers and other educators should know, for example, that the range of measured mental ability in grade I is x-years above and below the mean. For without such knowledge, sound and effective instructional planning and method are not possible, and classrooms would revert to the former lock-step methods. It is essential, also, that educators should know the ranges of abilities, both general and special, in subsequent grades; and they should know that the ranges of absolute differences increase with older age-groups until the maturity level is reached, as found under present conditions of mental development and with the use of present instruments of measurement. So far as teachers, counselors, and school administrators are concerned, in respect to individual differences, they are faced with an actual condition as it exists at a particular time, and for which they must make the most effective educational provisions possible. For such provisions, knowledge of the pertinent facts of variations in respect to general ability, reading rate and comprehension, spelling, manipulation of arithmetic processes and arithmetical problem-solving, emotional and social maturity, physical maturity, etc. are necessary.

Unless teachers, counselors, and educators are to accept the reported data of human variations as fixed and predetermined 'by nature,' resulting in a rigid educational and social outlook, they must know also the theory that has evolved from the many years of research. For as Pasteur said, " . . . without theory, practice is but routine, born of habit. Only theory can bring forth and develop the spirit of invention." Without an understanding of theoretical interpretations of individual differences, teachers can be only rule-of-thumb technicians; and as such they cannot have the necessary insights into the behavior and learning of their pupils for the fostering of optimal development. Through an understanding of research and emerging theory, teachers and administrators can best appreciate and develop the practical applications.

There is a tendency in some quarters to depreciate, if not to scorn, the work and contributions of theorists. Just as a reminder, therefore, we may name, for example, the following men of the modern period who were or are primarily 'theorists,' but who have made major contributions to modern educational practice: John Dewey, C. H. Judd, Alfred Binet, J. McK. Cattell, E. L. Thorndike, C. Spearman, L. M. Terman, W. F. Dearborn, B. H. Bode, W. C. Bagley, W. H. Kilpatrick, and the Gestalt psychologists (notably Lewin, Wertheimer, Koehler, Koffka, and R. M. Ogden).

II

Specifically, then, in what areas should educators acquire scientific knowledge and theory, as related to the subject of individual differences? Without implying an order of relative importance, the following topics should be dealt with.

Human genetics: not only to understand the biological mechanism of human inheritance, but in order also to understand the reasons for and the possibilities of genetic variations within any family, as between parents and siblings and between siblings themselves; to understand that genetically a child is not necessarily fated to be 'like' his parents or siblings, and that an individual's genetic potentialities cannot be determined from general laws and general trends.

Intrafamily correspondence and differences: to follow directly the study of genetic principles and to demonstrate those principles, as well as to show the influences of environmental factors. Under this topic would be included the familiar subjects of parent-child and sibling relationships, studies of fraternal and identical twins, and of foster children. From these materials, the principle of interrelationship and interdependence of nature and nurture clearly emerges.

Differences due to racial or national origins: to understand their relations to cultural factors (social, economic, educational) as these differ among nations and among racial and national groups within one nation (caste and class membership). Knowing the facts of individual differences in respect to racial membership or national origin of parents is, of course, necessary; but knowing the causal relationships between psychological traits and social forces is essential if we are concerned with implementing a democratic philosophy of education. For, without knowing these causal relationships, how can we hope to understand the reasons for existing conditions and the means whereby we, as social scientists, may remedy present defects and deficiencies?

Differences due to socio-economic factors: to understand these factors and their relationships to parental occupation, family income, type of community; and all that these conditions imply with respect to educational opportunity, developmental resources in the home, educational motivation and tradition. The study of this aspect of the subject should include among others, the investigations made in remote and psychologically impoverished areas (e.g., isolated mountain children, canal-boat children); studies of children in rural and urban areas; of the influence of nursery-school attendance upon measured abilities, at one end of the educational structure, and of high-school and college attendance at the other end. Under this caption, differences in abilities due to sex membership might also be included, since they are largely culturally determined, although in a much smaller degree biological factors are also operative, as in the more rapid early rate of development of girls.

The nature and causes of so-called special abilities and disabilities: to clarify the facts regarding the distribution of abilities, especially in music and graphic arts, and of mechanical skills. But it is equally important that individual differences in these abilities be viewed in the light of the general lack of educational opportunities and emphasis in these arts and skills in the lives of the vast majority of children, as compared with the heavy educational emphasis placed upon development of verbal and numerical abilities. By contrast, to demonstrate that these abilities are not merely the unfolding and expression of 'heredity,' the lives and training of distinguished persons in these fields should be studied in order that nature and nurture in these abilities may be seen in proper perspective. Nor, in this connection, may we neglect the influences of the social value or lack of value associated with a field of study and which consequently affects purpose and motivation.

Differences in mental organization: to demonstrate that while many abilities are significantly or highly correlated, they are by no means perfectly correlated; that individual profiles of abilities differ, and that effectiveness of mental operations in complex situations depends upon the organization of the several psychological functions.

The psychology of the more extreme deviates: namely, the mentally deficient, the gifted, the non-reader, the physically handicapped, and the sensory handicapped. While most of these groups are regarded as the more extreme deviates on a continuum, their remoteness from the modal area is so great as to create qualitative differences which cause special problems of psychological adjustment and educational practice.

The individual as a whole: to demonstrate the current concept that an individual at all times functions as a whole. Thus the study of personality development and personality factors is essential for a fuller appreciation of individual abilities and their operations. This aspect would encompass the usual range of non-intellective personality traits, including values, attitudes, and interests in their relationships to the general culture and the subcultures within which everyone lives and develops.

Before a teacher or prospective teacher could undertake the study of the foregoing aspects of the comprehensive subject of individual differences, preparation in several other basic fields is indicated: namely developmental psychology, the psychology of motivation (what is commonly called 'needs' or 'drives'), psychological and educational testing, and elementary statistical reasoning. Without adequate preparation in these basic subjects, the educator will be seriously handicapped in evaluating and interpreting the significance of the factual materials in the study of individual differences.

If all these suggested fields of study appear to be a large order, then the only answer to be given is that the profession of teaching is a complex and significant one; and preparation for it, therefore, needs to be complex and intensive.

A democracy is one of the most exacting forms of society in terms of its requirements in individual responsibility. That being the case, the teaching profession is the keystone of the arch, especially at the elementary- and secondary-school levels; because it is here, in these phases of life, that the foundations and patterns of living are developed. Thus, in addition to practical training and experience, the professional preparation of teachers demands the kind of scientific and theoretical study that is the distinguishing characteristic between the professional and the technician.

Equipped with the information and theory concerning individual differences—both inter- and intra-individual—educators will be in a much more secure position to evaluate current educational practices and to develop others. They will be in a sounder position to judge the results of psychological procedures which so profoundly affect educational practice and philosophy. For example, they will be able better to interpret the meaning of an IQ and will not be unduly influenced by uncritical enthusiasts of mental testing, nor by the misinformed who reject the IQ because it is not as invariable as the movement of celestial bodies. They will be better able to evaluate the arguments for and against ability grouping. They will be freed from the tendency to pass judgment upon a child or adolescent on the basis of his skin color, parents' national origin, parental occupation, sibling's ability, or sex membership. They will have a clearer conception of the concept of 'the child as a whole,' often so vaguely employed. They will also be aware of the fact that while the range of individual variations in respect to any trait is great, and while every person as an integrated whole is unique, a very large portion of the members of a culture or subculture have much in common, in regard to mental abilities and basic needs; and, that within a given culture, they have much in common regarding their modes of living and acting. This communality of abilities, needs, and traits is especially apparent in the large group of persons who constitute the 'average group'; that is, approximately the sixty per cent who are at or relatively close to the central tendency of a distribution.

No one in any profession can be prepared, in a professional school, to meet all contingencies. All professions are and should be constantly in process of development. In the several fields of psychology which are applied to educational problems and procedures, we are dependent upon teachers and other professional personnel in the schools not only for the effective application of psychological facts and principles to the educative process, but we are also dependent upon them for critical and insightful observation and assessment of these facts and principles. When they have a full appreciation of the work of the scientists and experimentalists and theorists, they are also most effective as professional practitioners.

BOOK REVIEWS

HAROLD E. JONES. *Motor Performance and Growth.* Berkeley: University of California Press, 1949, pp. 181.

This investigation is concerned with the development of muscular strength as revealed by semiannual dynamometric measurements of eighty-nine boys and eighty-seven girls from age eleven to seventeen and one-half years. Although the author considers it of little significance, it should be noted that the occupational distribution of the sample is not representative of the general population. There was a marked excess of white-collar workers and skilled workmen, and almost none from the semiskilled and unskilled groups. Interpretations, therefore, probably should be confined to the present sample rather than extended to the general population.

The reliability coefficients, by a split-test method or by test vs. retest after a week, are high, i.e., ranging from .901 to .964. Constancy of performance over an interval of six months or longer, however, tends to decrease considerably, especially for the girls. Thus, over a five-year period, the coefficients are approximately .60. Maturity factors appear to be involved.

Correlational analyses of various kinds were made. Space permits citation of only a few of the findings: For boys the correlation of grip strength with various gross motor abilities averaged about .50, but was very low for girls. Strength is relatively independent of socio-economic factors for the group. Among boys there is a slight positive relation between strength and social prestige. Examination of various types of growth curves and case summaries furnished additional information.

Educational implications are emphasized. New tasks "include improving the general standards of physical performance, helping children with physical deficiencies to approach these standards, and helping them to make secure and realistic adjustments to deficiencies which cannot be overcome."

It is probably best to receive many of these findings as tentative and suggestive rather than firmly established until additional studies are made with more representative samples.

MILES A. TINKER

University of Minnesota

ERNEST GLEN WEVER. *Theory of Hearing.* New York: John Wiley & Sons, Inc., 1949, pp. 441.

The book begins with a presentation of all important classical theories on hearing from the beginning of the Seventeenth Century to the present. At the time of Dr. Wever's first interest in the theory of hearing, about two decades ago, the two dominant theories—the place and frequency theories—were diametrically opposed to each other. In collaboration with C. W. Bray, in 1930, he discovered the electrical potentials of the ear, and his tests quickly showed that old and new evidence could be reconciled with both conflicting theories. Continued research by electro-physiological techniques on the ear's behavior forms the basis of his fusion into a compromise position of the two theories into a new one—the volley theory. Each of the older theories rests upon substantial evidence, but neither by itself is sufficient. Naturally, many problems remain, and further extended investigation is needed. What happens to the peripheral patterns as they are transmitted centrally is not yet known in detail. Also unsolved is the physical basis of cochlear differentiation. However, further pursuit of the problems can be carried on more comfortably because it is no longer necessary to load the whole burden of theory upon mechanical conditions, and the frequency variable can be allowed to take its share.

This is, indeed, a thorough and profound work by Dr. Wever, who is Professor of Psychology at Princeton University. There is a splendid frontispiece illustration of the anatomy of the human ear by the late Max Brödel, of Hopkins. A twenty-two-page bibliography of important references, four pages of definitions and symbols, and a fourteen-page index complete the volume.

LOUIS CHESLOCK

Peabody Institute, Baltimore

HUGH WOODWORTH. *The Nature and Technique of Understanding: Some Fundamentals of Semantics.* Vancouver, B.C.: The Wrigley Printing Company, Ltd., 1949, pp. 142.

Most of this book is devoted to the psychology of consciousness—the cognitive aspects particularly. The reader who expects to find a great deal about techniques of understanding will be disappointed. The author's avowed aim is to discuss meaning and understanding and the technique for attaining them. After devoting the first half of the book to a discussion of topics ranging from consciousness and relation of mind and body to feelings, sensations and verbal and extended meaning, Woodworth concludes that understanding is a high degree of extended meaning. Extended meaning is dependent upon past experience and knowledge, and can be realized only by dwelling on a word or statement and by exerting effort, whereas verbal meaning is the quick, ready and automatic minimum meaning evoked.

The author then classifies experience into dynamic and induced experience. Dynamic experience is defined as consciousness produced by the body in action and by the effect of the external world upon the receptor organs, whereas induced experience is the stream of consciousness in reverie and thought. The importance of the kinaesthetic sense in dynamic experience is stressed.

Woodworth finds it necessary to remind the reader in the opening statement of Chapter 8 of this ten-chapter book that the main topic is still understanding, although he had to wander far afield in laying the foundation for the discussion of this topic. This remark gives one some idea of the devious path by which the author arrives at the treatment of his main topic. The author's theme is 'to understand is to experience.' Induced experience which approximates or almost duplicates the dynamic experience of a thing is the essence of understanding a thing. Understanding of things of which one has had no similar dynamic experience is bound to be weak. The rôle of feelings, freelings (the author's term for feelings accompanying sensations), and sensations in experience and understanding is stressed. As an illustration of how to employ these in understanding, the author suggests that one can understand Ohm's Law more fully by kinaesthetically experiencing the relationships of the variables; or, preferably, by identifying oneself with the electric current moving in the wire and actually experiencing the circuit.

The author makes a distinction between meaning and understanding and realization. Meaning refers to the relatively passive associations aroused by a stimulus, whereas understanding and realization involve the organization of the associations so as to duplicate the experience which is the referent of the stimulus.

On page 124 the author finally sees fit to discuss techniques for aiding understanding. After examining the factors which enter into or determine understanding; namely, effectiveness of words, range of dynamic experience possessed, inhibitory control over associations set off, and degree to which associations can be restructured to yield induced experience duplicating or paralleling the indicated dynamic experience, Woodworth decides that the last two factors can be improved and that a wide background of experience can be cultivated. He suggests that a statement can be more fully understood by dramatization—by acting it out. He exhorts all who would increase their understanding to cultivate the dramatist in themselves. He recommends absorbing things slowly, reflecting upon them, and attending to one's sensations and emotions as one contemplates. In the last chapter Woodworth offers an hypothesis, which seems dubious to the reviewer, on the deep rôle of feelings in understanding.

The reader of this book who possesses an elementary knowledge of psychology will put it down without having found a single new idea on how to improve understanding. Possibly some notions which he has had for a long time will be revitalized for a time after reading it. He will have enjoyed reading a well-written book on some theoretical topics related to understanding, but couched in a somewhat idiosyncratic vocabulary. The author demonstrates his own skill in communicating ideas to the reader, but fails to give the reader a technique for improving his understanding.

Not a single reference is made to any authority in the fields of semantics, psychology, or any other field of learning, in the entire book. No bibliography and no reference to any article or book appear in this work. Although it is apparent that the author has some acquaintance with the literature of psychology, he writes as though he were unaware of its existence. He gives the impression of writing a personal treatment of his subject, independent of the body of knowledge available; yet examination of the content of this book shows that this is not the case.

The book is recommended to those who would like to re-emphasize in their thinking the psychology of cognition. To those unacquainted with the field of psychology, this book may seem provocative. PHILIP M. KITAY

Adelphi College

HERSCHELL ALT (Chairman) AND OTHERS, *Children Absent from School.* New York: Citizens' Committee on Children of New York City, 1949, pp. 116.

This report, which began as a study of children with records of frequent and unexplained non-attendance, is a contribution toward the establishment of a comprehensive child-helping program in the New York City schools. The three parts of the report are concerned with the nature of the problem of non-attendance, the bureau of attendance in relation to the non-attendance problem, and the proposed program of adjustment services for the school system.

The study revealed that most of the children referred to truant officers turned out to be lawful absentees, i.e., only fifteen per cent were truants. And for these truants, the unlawful absence was found to be only a symptom of adjustment problems at home, at school or within the child's personality. Underlying causes are not removed by treating symptoms. These children need a *kind of help that* the bureau of attendance is not qualified to give, i.e., to help maladjusted children. Since truancy is one among many adjustment problems in the school system, the report rejects any patchwork improvement of individual bureaus and asks for a comprehensive plan of adjustment services. In such a plan, emphasis is placed upon the rôle of the classroom teacher who must initially recognize trouble and find help. As a new service within the Division of Child Welfare, a school counselor is proposed for all schools as a skilled consultant available to the teacher. Another special bureau in the same Division should be empowered to provide help to child or parent when needed. Several other recommendations deal with special services and organization of the school adjustment program.

This excellent report, although concerned with a particular school system, has important implications for other systems.

MILES A. TINKER

University of Minnesota

FREDERICK MOSTELLER, HERBERT HYMAN, PHILLIP J. MCCARTHY, ELI S. MARKS, AND DAVID B. TRUMAN WITH THE COLLABORATION OF LEONARD W. DOOB, DUNCAN MACRAE JR., FREDERICK F. STEPHAN, SAMUEL A. STOUFFER AND S. S. WILKS. *The Pre-election Polls of 1948: Report to the Committee on Analysis of Pre-election Polls and Forecasts.* New York: Social Science Research Council, 1949, pp. 416. $2.50 (paper)

Much that is incisive and much that is guff has been written since November, 1948, concerning the prediction failures of the various pre-election polls. Educators, social scientists and students seeking an unbiased and comprehensive account of sources of error in the forecasts can find it in this bulletin. There is a minimum of diatribe and a maximum of technical information concerning polling practices. Much of this information is presented for the first time in the literature.

Immediately following the election, the Social Science Research Council decided to sponsor an authoritative factual inquiry into the polls' difficulties. This centralized research, it was hoped, would focus attention on the most pertinent specific issues and hold down extended controversy over minor points. It would avoid duplication of inquiry by many committees with a variety of sponsors and it would marshal some of the best talent in the nation as inquiry participants.

A Committee on Analysis of Pre-election Polls and Forecasts and a technical staff were assembled by mid-November, 1948, with a roster resembling a Who's Who in social psychology and survey research. Both national and local polling organizations opened their files completely to these investigators. As a result, the Committee was able to release a summary report by late December, 1948, based on the detailed findings of the technical staff. The detailed staff findings comprise the bulk of this bulletin, and the original committee report is included in the appendix.

Three early chapters deal with the objectives and limitations of the bulletin, with historical matters, and with a presentation of state-by-state election results alongside polling predictions. Some data from city polls versus election results also are given.

Leonard Doob's section on the public presentation of polling results was of particular interest to the reviewer. Using the technique of communications content analysis, Doob is able to shed a great deal of light on the problem of whether the polling organizations mishandled their public relations and oversold themselves and their services.

Technical sections on measuring prediction error and on sampling techniques will be of great interest to practitioners and serious students of survey methods. A comparison of the efficiency of polling predictions to a mere 'persistence forecast' is particularly interesting. An excellent job is done in detailing quota and pin-point samples as used in 1948. The discussion of probability sampling, however, probably will not be sufficient to clarify this topic for the non-professional reader.

Questionnaires used by the polling agencies were deficient chiefly in their comprehensiveness of design. This is considered to be a reflection of the rudimentary state of theory in the polling field. Available data on interviewer effects is rather meagre and often indirect. Precautions taken by major polls to minimize this error source are explained.

Chapters on processing and adjustment of data, last minute swing, and the undecided respondent illustrate how the polling organizations met the sticky problem of making final forecasts from their raw data.

A challenging section on political behavior points out the long road which social science must yet travel in formulating adequate theory of the behavior of the voter.

An elaborate appendix section contains the Committee Report, a complete breakdown of the material on which the Crossley, Roper, and Gallup forecasts were made, and additional data from the Washington State Poll, the Michigan Survey Research Center area sample poll, and the *Washington* (D.C.) *Post* Plebiscite.

Despite the cautions regarding forecasting engendered by this bulletin, the reviewer does not hesitate to predict for it an important niche in the survey research literature. Educators offering courses in this area will certainly find it invaluable as a principal or supplementary text. ROBERT L. JONES

University of Minnesota

PUBLICATIONS RECEIVED

Clifford R. Adams. *Looking Ahead to Marriage.* Chicago: Science Research Associates, Inc., 1949, pp. 48. (paper) $0.60.

Frederick B. Agard and Harold B. Dunkel. *An Investigation of Second-language Teaching.* Boston: Ginn and Co., 1948, pp. 344. $2.75.

Audio-Visual Education Association of California. Setting Up Your Audio-Visual Education Program. Stanford, Calif.: Stanford University Press, 1949, pp. 34. (paper) $1.00.

Lester F. Beck. *Human Growth: The Story of How Life Begins and Goes On.* New York: Harcourt, Brace & Co., 1949, pp. 124. $2.00.

Bruno Bettelheim. *Love Is Not Enough. The Treatment of Emotionally Disturbed Children.* Glencoe, Illinois: The Free Press, 1950, pp. 386. $4.50.

Bridges between the School and the Community. New York: Board of Education of the City of New York, 1950, pp. 79. (paper)

Thomas H. Briggs, J. Paul Leonard, and Joseph Justman. *Secondary Education.* Revised Edition. New York: The Macmillan Co., 1950, pp. 468. $4.00.

Steuart Henderson Britt. *Selected Readings in Social Psychology.* New York: Rinehart & Co., Inc., 1950, pp. 507. (paper) $2.00.

Oliver C. Carmichael, President. *The Carnegie Foundation for the Advancement of Teaching: Forty-fourth Annual Report.* New York: The Carnegie Foundation for the Advancement of Teaching, 1949, pp. 61. (paper)

Matthew N. Chappell. *In the Name of Common Sense: Worry and Its Control.* Revised Edition. New York: The Macmillan Co., 1949, pp. 172. $2.75.

Hector Chevigny and Sydell Braverman. *The Adjustment of the Blind.* New Haven: Yale University Press, 1950, pp. 320. $4.00.

Alois Cibulka. *All This Could Happen Only to an Engineer.* Houston, Tex. (3801 St. Bernard St.): The Author, 1950, pp. 237. $4.50.

Herbert S. Conrad. *Characteristics and Uses of Item-analysis Data.* Psychological Monographs: General and Applied, 62:8. Washington, D. C.: The American Psychological Association, 1949, pp. 49. (paper)

FRANCIS G. CORNELL, WILLIAM P. MCLURE, VAN MILLER, AND RAYMOND E. WOCHNER. *Financing Education in Efficient School Districts.* Urbana, Illinois: University of Illinois, College of Education, 1949, pp. 165. (paper)

ROGER COUSINET AND OTHERS. *Leçons de Pedagogie.* Paris, France: Presses Universitaires de France, 1950, pp. 263. (paper)

LORAINE ANSON DAHL. *Public School Audiometry: Principles and Methods.* Dansville, Illinois: The Interstate, 1949, pp. 290. $3.00.

JEROME EDWARD DOPPELT. *The Organization of Mental Abilities in the Age Range 13 to 17.* New York: Bureau of Publications, Teachers College, Columbia University, 1950, pp. 86. $2.10.

ROY M. DORCUS AND MARGARET HUBBARD JONES. *Handbook of Employee Selection.* New York: McGraw-Hill Book Co., 1950, pp. 349. $4.50.

PAUL L. DRESSEL AND OTHERS. *Comprehensive Examinations in a Program of General Education.* East Lansing, Mich.: Michigan State College Press, 1949, pp. 165. $4.00.

HAROLD B. DUNKEL. *Second-language Learning.* Boston: Ginn & Co., 1948, pp. 218. $2.75.

ALLEN L. EDWARDS. *Experimental Design in Psychological Research.* New York: Rinehart & Co., Inc., 1950, pp. 446. $5.00.

BERTRAM EPSTEIN. *Immediate and Retention Effects of Interpolated Rest Periods on Learning Performance.* New York: Bureau of Publications, Teachers College, Columbia University, 1949, pp. 77. $2.10.

FRANK S. FREEMAN. *Theory and Practice of Psychological Testing.* New York: Henry Holt & Co., 1950, pp. 518. $3.50.

G. L. FREEMAN AND E. K. TAYLOR. *How to Pick Leaders.* New York: Funk and Wagnalls Co., 1950, pp. 226. $3.50.

DOUGLAS H. FRYER AND EDWIN R. HENRY, Editors. *Handbook of Applied Psychology.* Two Volumes. New York: Rinehart & Co., Inc., 1950, pp. 842. $12.50.

Fundamental Education: A Description and Programme. Paris, France (19, Avenue Kleber): UNESCO, pp. 85. (paper)

J. Roswell Gallagher. *You and Your Health.* Chicago: Science Research Associates, 1950, pp. 48. (paper) $0.60.

Regina M. Goff. *Problems and Emotional Difficulties of Negro Children.* New York: Bureau of Publications, Teachers College, Columbia University, 1949, pp. 93. $2.10.

J. P. Guilford. *Fundamental Statistics in Psychology and Education.* New Second Edition. New York: McGraw-Hill Book Co., 1950, pp. 633. $5.00.

J. P. Guilford and William B. Michael. *The Prediction of Categories from Measurements: With Applications to Personnel Selection and Clinical Prognosis.* Beverly Hills, Calif.: Sheridan Supply Co., 1949, pp. 55. (paper) $1.48.

Ward C. Halstead, Editor. *Brain and Behavior: A Symposium.* Comparative Psychology Monographs. Vol. 20., No. 1. Berkeley, Calif.: University of California Press, 1950, pp. 94. (paper) $2.50.

Melvin S. Hattwick. *How To Use Psychology for Better Advertising.* New York: Prentice-Hall, Inc., 1950, pp. 376. $5.75.

D. O. Hebb. *Organization of Behavior: A Neuropsychological Theory.* New York: John Wiley and Sons., Inc., 1949, pp. 335. $4.00.

Nelson B. Henry, Editor. *The Forty-ninth Yearbook of the National Society for the Study of Education.* Part I: Learning and Instruction, pp. 352. (paper) $2.75. Part II: The Education of Exceptional Children, pp. 350 #lxii. (paper) $2.75. Chicago: The University of Chicago Press, 1950.

Wilson H. Ivins, William H. Fox, and David Segel. *A Study of a Secondary-school Program in Light of Characteristics and Needs of Youth.* Bloomington, Ind.: Indiana University, 1949, pp. 69. (paper)

Charles Douglas Keet. *Two Verbal Techniques in a Miniature Counseling Situation.* Psychological Monographs: General and Applied, 62:7. Washington, D. C.: The American Psychological Association, 1949, pp. 55. (paper)

Carney Landis and M. Marjorie Bolles. *Textbook of Abnormal Psychology.* Revised Edition. New York: The Macmillan Co., 1950, pp. 634. $5.00.

George Lawton. *How To Be Happy though Young.* New York: The Vanguard Press, Inc., 1949, pp. 300. $3.00.

WILLIAM C. MENNINGER. *Enjoying Leisure Time.* Chicago: Science Research Associates, Inc., 1950, pp. 48. (paper) $0.60.

MARY ALICE MITCHELL. *The Relationship of Reading to the Social Acceptability of Sixth-grade Children.* New York: Bureau of Publications, Teachers College, Columbia University, 1949, pp. 59. $2.10.

NORMAN L. MUNN. *Handbook of Psychological Research on the Rat. An Introduction to Animal Psychology.* Boston: Houghton Mifflin Co., 1950, pp. 598. $7.50.

CHARLES E. O'TOOLE AND OTHERS. *Manual of Procedures in Individual and Group Guidance in Junior High Schools.* New York: Board of Education of New York City, 1949, pp. 120. (paper)

E. H. PORTER, JR. *An Introduction to Therapeutic Counseling.* Boston: Houghton Mifflin Co., 1950, pp. 223. $2.75.

STANLEY D. PORTEUS. *The Porteus Maze Test and Intelligence.* Palo Alto, Calif.: Pacific Books, 1950, pp. 194. $4.00.

JOHN W. M. ROTHNEY AND BERT A. ROENS. *Guidance of American Youth.* Cambridge, Mass.: Harvard University Press, 1950, pp. 269. $5.00.

CHARLES E. SKINNER, Editor. *Elementary Educational Psychology.* Second Edition. New York: Prentice-Hall, Inc., 1950, pp. 592. $3.75.

ELIZABETH STADTLANDER. *Planning to Study Efficiently.* St. Louis (122 N. 7th St.): Educational Publishers, Inc., 1950, pp. 116. (paper) $2.25.

WALTER W. STEWART, Chairman. *General Education Board Annual Report for 1949.* New York: General Education Board, 1950, pp. 90. (paper)

C. P. STONE, Editor. *Annual Review of Psychology.* Volume I: 1950. Stanford, Calif.: Annual Reviews, Inc., 1950, pp. 330. $6.00.

ESTHER J. SWENSON, G. LESTER ANDERSON, AND CHALMERS L. STACEY. *Learning Theory in School Situations.* Minneapolis: The University of Minnesota Press, 1949, pp. 103. (paper)

DONALD R. TAFT. *Criminology: A Cultural Interpretation.* Revised Edition. New York: The Macmillan Co., 1950, pp. 704. $5.50.

WILLIAM CLARK TROW. *Educational Psychology.* Second Edition. Boston: Houghton Mifflin Co., 1950, pp. 761. $4.00.

The United States and the Soviet Union: Some Quaker Proposals for Peace. New Haven: Yale University Press, 1949, pp. 39. (paper) $0.25 (through American Friends Service Committee, 20 So. 12th St., Philadelphia).

P. F. VALENTINE, Editor. *The American College.* New York: Philosophical Library, 1949, pp. 575. $10.00.

PHILIP E. VERNON AND JOHN B. PARRY. *Personnel Selection in the British Forces.* London: University of London Press, (The British Book Centre, Inc., 122 E. 55th St., New York City), 1949, pp. 324. $4.50.

HARVEY S. WHISTLER AND LOUIS P. THORPE. *Musical Aptitude Test* (Series A.) Hollywood, Calif.: California Test Bureau, 1949, pp. 23. (paper) $3.00.

HUGH WOODWORTH. *The Nature and Technique of Understanding: Some Fundamentals of Semantics.* Vancouver, Canada: The Wrigley Printing Co., Ltd., 1949, pp. 142. $4.00.

PAUL A. ZAHL, Editor. *Blindness: Modern Approaches to the Unseen Environment.* Princeton, N. J.: Princeton University Press, 1950, pp. 576. $7.50.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 41 November, 1950 Number 7

RELIABILITY, HOMOGENEITY AND NUMBER OF CHOICES

N. L. GAGE and DORA E. DAMRIN*
Bureau of Research and Service College of Education,
University of Illinois

Reliability data are indispensable to any full evaluation of test validity, either for practical purposes or for theorizing about true relationships between traits. The traits themselves often depend, for evidence of their existence, on the demonstration that they are reliably observable.

But theorizing about test reliability often goes on in terms of mathematical manipulations of statistical concepts that are somewhat divorced from reality. Theorists develop assumptions and constructs to simplify their formulas but in doing so may disregard the psychological data to which their formulas will be applied. This creates the place for empirical studies. However pat the derivations, we still like to see how the formulas work when applied to typical data. Such investigations can tell us how close to reality the assumptions and definitions have stayed, or how much practical difference their being violated actually makes.

SINGLE-TRIAL ESTIMATES OF RELIABILITY

Our first major interest was in comparing those estimates of reliability that can be made from single trials. Of the methods, we shall consider only the Spearman-Brown corrected split-half coefficient (S-B split-half), Guttman's L_4, and the Kuder-Richardson Cases III and IV (K-R III and K-R IV). Other formulas, such as Rulon's,[17] Guttman's L_3 and Jackson's sensitivity (gamma)[15] all reduce to algebraic identities with, or mathematical functions of one of these (e.g., gamma = $\sqrt{\frac{r}{1-r}}$); hence, they are omitted here. Loevinger's estimate of homogeneity[12] will also be considered, although it is not intended as an estimate of reliability in the usual sense. We shall examine below the definitions of equivalence involved in each of these methods.

---

* We have been helped by many criticisms and suggestions from Lee J. Cronbach; however, full responsibility for interpretations and conclusions rests with us.

Spearman-Brown corrected split-half coefficients.—Here the single test is split into halves (usually odd- and even-numbered items) and scores on the halves are correlated. The halves are assumed to be equivalent both with one another and with the halves of a hypothetical equivalent form as follows:

$$\sigma_a = \sigma_b = \sigma_A = \sigma_B;\ r_{ab} = r_{aB} = r_{aA} = r_{bA} = r_{bB} = r_{AB}$$

where a and b are halves of the actual test, A and B are halves of a hypothetical equivalent form.

Using this definition of equivalence we apply the Spearman-Brown formula to estimate the reliability of the whole test, $a + b$, as follows:

$$r_{(a+b)(A+B)} = \frac{2r_{ab}}{1 + r_{ab}}$$

Guttman's L_4.—This formula, according to its author, allows us to "dispense with assumptions of equivalence,"[5] (p. 275). Cronbach[2] has shown, however, it can be derived by redefining equivalence so that $\sigma_{a+b} = \sigma_{A+B}$ and $r_{aA}\sigma_a\sigma_A = r_{aB}\sigma_a\sigma_B = r_{Ab}\sigma_A\sigma_b = r_{bB}\sigma_b\sigma_B = r_{ab}\sigma_a\sigma_b$. This redefinition of equivalence of an actual test with a hypothetical one leads to the 'lower bound' formula:

$$L_4 = 2\left(1 - \frac{s_1^2 + s_2^2}{s_t^2}\right)$$

Guttman's crucial assumption is experimental independence of items within trials.

Kuder-Richardson Case III.—These writers[11] define equivalence of forms as interchangeability of items i and I, j and J, etc.; the members of each pair have the same difficulty ($p_i = p_I$, $p_j = p_J$, etc.) and are correlated to the extent of their reliabilities: $\frac{r_{iI}}{\sqrt{r_{ii}r_{II}}} = 1$. The inter-item correlations of one test are the same as those in the other ($r_{ij} = r_{IJ}$, etc.). Since straightforward substitution of these equalities, into the formula for correlation between the two equivalent forms, leads to a need for unobtainable quantities like r_{ii}, they make further assumptions to achieve a workable formula. Among these are (1) that "item and test measure the same thing" $\left(\frac{r_{it}}{\sqrt{r_{ii}r_{tt}}} = 1\right)$, (2) that all item intercorrelations are equal, thus yielding a matrix of inter-item correlations which has a rank of one ($r_{ab} = r_{ac} = \cdots = r_{n(n-1)}$), and (3) that item variances are equal ($p_aq_a = p_bq_b$, etc.). The resulting "formula 20" is:

$$r_{tt} = \frac{n}{n-1} \cdot \frac{\sigma_t^2 - \Sigma pq}{\sigma_t^2}$$

where n = number of items
σ_t^2 = variance of test
p = proportion passing each item
$q = 1 - p$

As Jackson and Ferguson have pointed out,[9] (p. 76) the assumption of equal inter-item correlations and equal item standard deviations imposes a further requirement, unstated by Kuder and Richardson, that the items be equal or at most of two levels in difficulty. Loevinger has concluded that the method therefore "applies only to a case of no importance," since complete satisfaction of these assumptions leads to "a test on which everyone scored either zero or perfect,"[12] (p. 11). The coefficient is defensible however since its authors point out that "if the assumptions are not met, the figures obtained are under-estimates,"[11] (p. 159).

Kuder-Richardson Case IV.—The formula in this case is

$$r_{tt} = \frac{n}{n-1} \cdot \frac{\sigma_t^2 - n\bar{p}\,\bar{q}}{\sigma_t^2}$$

$$\text{where } \bar{p} = \frac{\text{Mean of the test scores}}{n} \text{ and } \bar{q} = 1 - \bar{p}.$$

Here the assumptions are the same as for Case III with the addition that all items are explicitly assumed to have the same difficulty. Since this assumption is implicitly made in Case III, as mentioned above, this formula represents merely, as Loevinger puts it,[12] (p. 13), "carrying one step further the consequences of assuming items of identical difficulty." Estimates by K-R IV have some merit in that the economy of effort in computing them counterbalances the still greater underestimation obtained as compared with K-R III. But their assumption of equal difficulty implies that they will lead to relatively inaccurate estimates of reliability for tests intentionally made to have items with a wide range of difficulty.

The Concept of Homogeneity.—Loevinger[12] has designed a formula to express a property of tests which is somewhat different from that approached by the above formulas and which she calls 'homogeneity.' This is the degree to which all items of a test measure the same ability or complex of abilities for all individuals tested. A necessary (but not sufficient) condition of perfect homogeneity is that all persons succeeding with items at one level of difficulty should also succeed with all items at lower levels of difficulty. Similarly "when the items of a perfectly homogeneous test are arranged in order of increasing difficulty, every individual will pass all items up to a certain point and fail all subsequent items,"[12] (p. 28). Subsequently[13] Loevinger pointed out that these requirements apply in this form only to what she has denoted 'cumulative' as against 'differential' homogeneous tests.

Characteristics of tests which are closely related to Loevinger's 'homogeneity' have frequently been discussed by others using such terms as coherence,[10] and unidimensionality or scalability.[6] Perhaps the first consideration of this characteristic was by Walker,[19] who in 1931 described a property of tests which he called 'higgledy-piggledyness.' To measure this property, Walker developed a 'coefficient of hig'[20] which he later[21] found of dubious merit.

In 1937, Newcomb verbally described the same characteristic of cumulative 'homogeneous' tests as applying to attitude scales: "No scale can really be called a scale unless one can tell from a given attitude that an individual will maintain *every* attitude falling to the right or to the left of that point (depending on how the scale is constructed),"[16] (p. 897).

In 1941, Ferguson,[4] (p. 54) stated that "at present no convenient quantitative measure is available for estimating the divergence of an obtained answer-pattern matrix from a theoretically unique matrix." Later,[4] (p. 66) he describes a measure of homogeneity which is equivalent, in Loevinger's terms, to the variance of the test divided by the variance of a perfectly homogeneous test with the same item difficulties.

Loevinger's index is an attempt at rational quantification of the concept. We shall examine its relationship to the concepts of 'reliability' embodied in the other operational definitions considered here. Her 'homogeneity' purports to be not another type of 'reliability' coefficient but a partial alternative to it. Her index relegates to error variance all of the following: group factor variance, specific factor variance, and errors of measurement of items. This means that a perfectly homogeneous test is one whose entire variance is entirely attributable either to a single general factor, or to "an approximately constantly weighted sum of factors,"[13] (p. 522).

Loevinger points out that her index of homogeneity is only one of many possible ways of estimating this property of tests. Its sampling properties are as yet unknown. We have studied it here because no empirical studies of the index have yet appeared. Our assumption is that the interpretability of the statistic will be enhanced by a knowledge of how it varies with number of choices and test length, and of how its size compares with the coefficients yielded by single-trial estimates of 'reliability.'

Since the estimate of homogeneity,

$$\text{est } H_t = \frac{V_x - V_{het}}{V_{hom} - V_{het}}$$

where V_x = Variance of the test,

V_{het} = Variance of a perfectly heterogeneous test with the given item difficulties,

V_{hom} = Variance of a perfectly homogeneous test with the given item difficulties,

it should be noted that it differs from K-R III by the omission of the $\frac{n}{n-1}$ term, and in the denominator. In Loevinger's terms,

$$K\text{-}R\ III = \frac{n}{n-1} \cdot \frac{V_x - V_{het}}{V_x}$$

TEST RELIABILITY AND NUMBER OF CHOICES

The second major problem at which this study is aimed is the relationship between single-trial estimates of reliability and number of choices per test item. It has long been realized that some of the error variance in multiple-choice tests of ability arises from chance success in choosing the correct alternative from those furnished in each test item. It should follow that increased test reliability will result from an increase in the number of functioning (in the sense of increasing the *r* between correct choice and total score) choices per test item.

The Spearman-Brown Formula.—This hypothesis has already been investigated in a series of studies by Remmers and his co-workers. Three of these studies[1,3,7] are pertinent here since they involved especially-designed experiments using ability tests in which each item is scored zero or one. In these studies Remmers investigated whether the changed reliability could be predicted by means of the Spearman-Brown formula, with change in test length conceived as the ratio of new number of choices per item to old.

In their studies it is evident that the sets of four reliability coefficients, although they did not differ significantly (in terms of critical ratios) from the Spearman-Brown predictions, did not in all cases follow a curve similar to that obtained by predicting with the Spearman-Brown formula from the two-choice form. This is especially true of House's data, in which the four-choice form was much more reliable than the Spearman-Brown formula predicted and the five-choice form showed almost no increase in reliability over the four-choice form. These discrepancies indicate the desirability of further investigation.

Two refinements of experimental procedure could be introduced to eliminate certain possible sources of the observed discrepancies between actual and predicted 'reliability' coefficients. The first is controlled rather than random elimination of misleads to convert five-choice items into four-, three-, and two-choice items. The second is administration of a 'control' test that would be the same for all groups, thus making possible empirical verification of the groups' comparability on the experimental forms and adjustment of the obtained coefficients for differences in the 'range of talent' of the groups. These refinements are further described below.

Lord's Formula.—Since the Spearman-Brown formula was originally developed to express the correlation between sums and differences, one cannot assume its relation to the correlation between forms of a test differing only in number of choices per item. Lord[14] objected to its use in this way on the ground that items cannot be made perfectly reliable simply by adding choices, and the Spearman-Brown formula does not take into account the difficulty of the test.

Lord has sought to avoid these difficulties by deriving an adaptation of the Spearman-Brown formula. Assumed in his derivation are (1) equal inter-item correlations (four-fold point r's), (2) equal item difficulties, (3) equal plausibility of choices within each item to individuals not knowing the correct answer, (4) answers to all items by all individuals, and (5) values of 0 or 1 for the choices in each item.

It can be objected that the assumptions of neither the Lord nor the Spearman-Brown formula will ever be met by actual test situations. It is nonetheless relevant to the evaluation of them, and to the theory underlying their development, to determine which more closely predicts the changes in reliability found empirically when changes in number of choices are introduced into a given test.

PROCEDURE

Our procedure in investigating the two major questions outlined above may be described in terms of (1) the test used, (2) the subjects tested, and (3) the administration of the tests.

The Test and Its Forms.—In choosing the test to be used, we had several considerations in mind. The test should be fairly homogeneous in content so that application of Kuder-Richardson formulas would be not altogether unjustified. It should be well-refined in terms of the discriminatory power and difficulty of its items and their misleads, so that we could have some assurance that all items and misleads were functioning. There should be adequate data on the difficulty and discriminatory power of the items and misleads so that some empirical basis could be used in the construction of the various forms. It should be a power test rather than a speed test since these formulas render spurious estimates of r_{tt} for speed tests. It should be an ability test rather than a non-intellective test so that the assumptions of certain of the formulas to be applied would be met. It should be applicable over a fairly wide range of ability so that students in both high-school and college classes could be tested. Finally, the scores obtained should have some promise of usefulness to the coöperating teachers and students to repay them for their time and effort.

The test selected as meeting all these requirements was the Ohio State University Psychological Test, Form 21, which has been developed and refined over many years by H. A. Toops and his collaborators. Of the one hundred fifty items only the first ninety, comprising Parts *1* and 2, were used. The first part consists of thirty same-opposite vocabulary questions illustrated by the following:

Little is the SAME AS 1 coarse 2 small 3 prodigious
4 immense 5 feeble

The second part consists of sixty verbal analogies of the following kind:

boy: boys: : man:
1 girls 2 men 3 man's 4 men's 5 gentlemen

These 90 items were arranged in order of difficulty on the basis of item analysis data.[11,*] The odd-numbered items, after this rearrangement, were designated 'control' items; they were left unchanged and were administered in the same form to all subjects. These 'control' items were intended to provide a means of verifying the comparability of the groups of subjects that were tested with 'experimental' items described below.

The even-numbered items were treated as the 'experimental' test by having their number of choices varied. Four forms of this test were made as follows: Form 5 consisted of the even-numbered items with five choices, i.e., intact, as in the original test developed by Toops. Form 4 consisted of the same items with one of the false choices eliminated; thus each of the Form 4 items had only four choices. Form 3 similarly had only three choices per item, i.e., two of the four original false choices were eliminated. Finally Form 2 had only two choices per item, three of the original four false choices having been eliminated.

The dropping of false choices to make Forms 4, 3, and 2 was done in terms of the item analysis data furnished by Toops.

---

[*] We are grateful to Professor Toops for his permission to use this test and for the item analysis data, based on 1,000 Ohio State University freshmen, which he supplied.

To make Form 4, the experimental items were arranged in order of difficulty. The items were then put into four groups, every fourth item going into the same group. The distributions of the item-test validity coefficients for these four groups were compared and found to be similar. From the first of these groups, the most popular mislead was eliminated; from the second group, the second most popular mislead was eliminated, etc.

To make Form 3, the same procedure was used except that only three groups of items, roughly matched on difficulty and validity, were formed. Of the remaining three misleads, the most plausible was dropped from the first group of items; the second most plausible, from the second group; etc. Form 2 was made by a repetition of this procedure after the items had been divided into two groups.

This procedure gave us some assurance that the forms would not differ sharply in respects other than the number of choices. The plausibility of the eliminated choices and hence of those remaining was controlled to produce a more representative manipulation of the items than chance alone might have yielded.

The Subjects.—Approximately one thousand high-school and college students were tested.* The high-school students were the entire eleventh and twelfth grade classes of three high schools, of which two were predominantly urban and the third mainly rural. The college students were for the most part sophomores and juniors enrolled in various sections of an introductory educational psychology course. Of those obtained, only the nine hundred seventeen answer sheets were used in subsequent analyses which had one and only one answer to every item.

Administration of the Tests.—The students were directed to proceed through the questions in the order given, to attempt every question, and to guess if necessary. These instructions were intended to eliminate individual differences in 'willingness to guess' as a possible non-intellectual factor affecting scores and, in effect, made the test more constant for all subjects. Furthermore, the 45-minute time limit proved sufficient to allow more than ninety per cent of the subjects to attempt all items. Since the incomplete answer sheets were eliminated from the statistical analysis, the results are based on what was essentially an unspeeded, power test for all subjects.

---

* For permission to test their students, we are grateful to Mr. James K. Felts, Monticello H.S.; Miss Sally Fisher, Urbana H.S.; Miss Vera Kaden, Champaign H.S.; Professors W. R. Dixon, Ray Simpson and Graham Pogue, University of Illinois.

The tests were distributed to the students in rotated order so that the first student in any room received Form 5, the second Form 4, etc. This mechanically rotated distribution was intended to yield random selection of students, and hence comparable groups, for each form. Scores on the 'control' items provided a means of ascertaining, and statistically adjusting for, such dissimilarity between these groups as still appeared.

RESULTS

Four scores (number right) were obtained for each student. Score 1 was based on the forty-five odd-numbered 'control' items. Score 2 was based on the forty-five even-numbered 'experimental' items, which differed from form to form in number of choices. Scores 3 and 4 were based on odd and even halves of the 'experimental' items, respectively, to provide data for computing those reliability estimates (corrected split-half and Guttman's L_4) which require splitting the test.

'Control' Test Results.—Table 1 shows the means and standard deviations of the 'control' scores of the four groups. Also given are the means and standard deviations of the scores on the 'experimental' items.

These figures show that the mean scores on the control items did not differ markedly from one group to another. Indeed, as the F-ratios of the analysis of variance indicate, the 'between groups' variance for the total group, far from being significantly large, is significantly smaller than chance would allow. This merely reflects the wide range of talent tested and the efficiency of the rotated distribution of forms in securing similar groups.

As is to be expected, the mean scores on the experimental forms decrease regularly (except for the college students on Form 5) from Form 2 to Form 5. The forms increased in difficulty (or decreased in susceptibility to chance success and to success through partial knowledge) as the number of choices increased. Since we wished in any case to make adjustments in the reliability estimates for such variations in 'range of talent,' we did not apply tests of significance to the differences among standard deviations of the four 'form-groups' on the control items.

TABLE 1.—MEANS AND STANDARD DEVIATIONS ON 'CONTROL' AND 'EXPERIMENTAL' ITEMS

Test Form	N	Control M	Control SD	Exper. M	Exper. SD	Results of Analysis of Variance Between Groups on Control Items
				Total Group		
2	241	18.32	10.20	30.06	7.55	
3	223	19.24	10.03	24.55	9.12	
4	228	18.33	10.52	21.23	10.49	F = .02, p > .05
5	225	18.56	10.25	19.23	10.40	
			High-school Juniors			
2	84	13.26	6.66	26.73	5.76	
3	80	13.05	6.35	20.62	7.33	
4	85	13.33	7.95	16.67	8.80	F = 1.81, p > .05
5	84	15.35	7.80	15.89	7.93	
			High-school Seniors			
2	96	15.48	7.89	28.60	6.96	
3	81	15.99	8.00	22.59	8.20	
4	82	16.46	9.50	19.68	10.07	F = .70, p > .05
5	87	14.97	7.97	15.79	8.95	
			College Students			
2	61	29.74	8.63	36.93	6.17	
3	62	27.87	9.60	32.16	7.71	
4	61	27.82	8.63	29.67	7.98	F = .70, p > .05
5	54	29.37	9.27	29.96	8.52	

Comparisons of 'Reliability' Formulas.—In Table 2 are given the various coefficients obtained on the 'experimental' test. On the left-hand side are the coefficients actually obtained; on the right are the coefficients adjusted to constant variability, or 'range of talent.' The standard deviation of the Form 2 group on the control items was used as the 'anchor' to which the reliabilities of the experimental forms taken by the other groups were adjusted. In comparing the adjusted coefficients shown in Table 2, we can see that (1) the Spearman-Brown corrected odd-even coefficients are all slightly larger than those obtained by Guttman's L_4, (2) the K-R III coefficients are all larger than those obtained by K-R IV, (3) the Spearman-Brown corrected odd-even coefficients are larger than those by K-R III except for Form 5, (4) the K-R III coefficients are not consistently

TABLE 2.—COEFFICIENTS OF 'RELIABILITY' FOR THE FOUR EXPERIMENTAL FORMS

Total Group

Form	Obtained				Adjusted for Variability on Control Test			
	Spearman-Brown Corrected Odd-even	Guttman L_4	Kuder-Richardson Case III	Kuder-Richardson Case IV	Spearman-Brown Corrected Odd-even	Guttman L_4	Kuder-Richardson Case III	Kuder-Richardson Case IV
2	.867	.858	.863	.844	.867	.858	.863	.844
3	.906	.901	.898	.886	.909	.904	.902	.890
4	.934	.929	.928	.918	.930	.925	.923	.913
5	.926	.923	.928	.918	.925	.922	.927	.917
High-school Juniors								
2	.797	.796	.737	.688	.797	.796	.737	.688
3	.854	.852	.831	.810	.867	.865	.840	.827
4	.901	.900	.898	.884	.859	.858	.855	.835
5	.855	.854	.870	.855	.801	.800	.822	.801
High-school Seniors								
2	.814	.814	.828	.803	.814	.814	.828	.803
3	.869	.867	.871	.852	.865	.863	.868	.848
4	.918	.911	.920	.911	.881	.871	.884	.871
5	.909	.907	.905	.892	.907	.905	.903	.890
College Students								
2	.859	.846	.865	.845	.859	.846	.865	.845
3	.899	.890	.886	.805	.875	.864	.859	.833
4	.911	.906	.886	.860	.911	.906	.886	.860
5	.903	.893	.905	.882	.888	.877	.891	.864

related to the L_4's, being slightly larger in two cases and slightly smaller in two.

The results shown in Table 2 do more than provide an empirical demonstration of known mathematical relationships between the several formulas. Of practical importance to the test-user in any field is the magnitude of the differences between various kinds of estimates of reliability. It will be noted that for a test of this type the 'underestimate' yielded by Guttman's L_4 is so slight as to be unimportant in practice. Likewise, the differences between the two types of K-R coefficients are, for most practical purposes, not of sufficient size to justify the extra labor involved in the computation of K-R III. The latter finding is especially interesting in view of the wide variation in the difficulty of these test items.

Although the estimates obtained by the split-half procedure are for the most part higher than those by the K-R formulas, this in no way establishes the superiority of this method. Such a finding is the result of chance, for the size of the coefficient obtained for any specific split is a function of the ratio of the sum of the half-test covariances to the inter-half covariances. If our split happens to be such as to maximize the within-half covariance terms and minimize those between halves, our coefficient will be low; if the reverse occurs the coefficient will be appreciably higher.

Comparison of Forms Differing in Number of Choices.—If the adjusted coefficients shown in Table 2 are compared from form to form, it is seen that they rise regularly from Form 2 (two-choice) to Form 4 for the total group. In two of the coefficients (S-B odd-even. and L_4) there is a slight drop in Form 5 from Form 4 while in the other two (K-R III and K-R IV) there is a slight continued rise from Form 4 to Form 5. In the subgroups by educational level, the trends are not so regular. Among the high-school juniors, the split-half coefficients (S-B odd-even and L_4) are highest for Form 3 while the K-R III and IV are highest for Form 4. Among the high-school seniors, all coefficients regularly increase from Form 2 to 5. Among the college students, the peak is hit at Form 4 for the split-half coefficients and at Form 5 for the K-R's.

How are these trends to be interpreted? In the first place, the general hypothesis that reliability increases as number of choices is increased tends to be supported by these data. When it is remembered that the K-R coefficients are unique estimates and depend in no way on judgment or chance in test-splitting, confidence in the hypothesis is further bolstered. Thus, the failure of the Form 5 split-half coefficients to show continued increase over Form 4 merely reflects 'bad luck' in splitting this test into equivalent halves.

Secondly, the level of ability of the group tested may be related to the number of choices at which maximum reliability is obtained. The fact that the highest K-R coefficients appear at Form 4 for high-school juniors but at Form 5 for both high-school seniors and college students indicates that the addition of choices may indeed lower reliability by increasing the difficulty of the test beyond the optimal point for the level of ability of the group.

Comparison of Lord and Spearman-Brown Formulas.—We have already described the two formulas that have been suggested for predicting change in test reliability with change in number of choices per test item. We shall evaluate them in terms of how closely their predictions accord with our adjusted K-R III coefficients.

Figure 1 shows the adjusted coefficients for all students by the K-R III formula, and those predicted from Forms 2, 3, and 4 by the Spearman-Brown and Lord formulas.

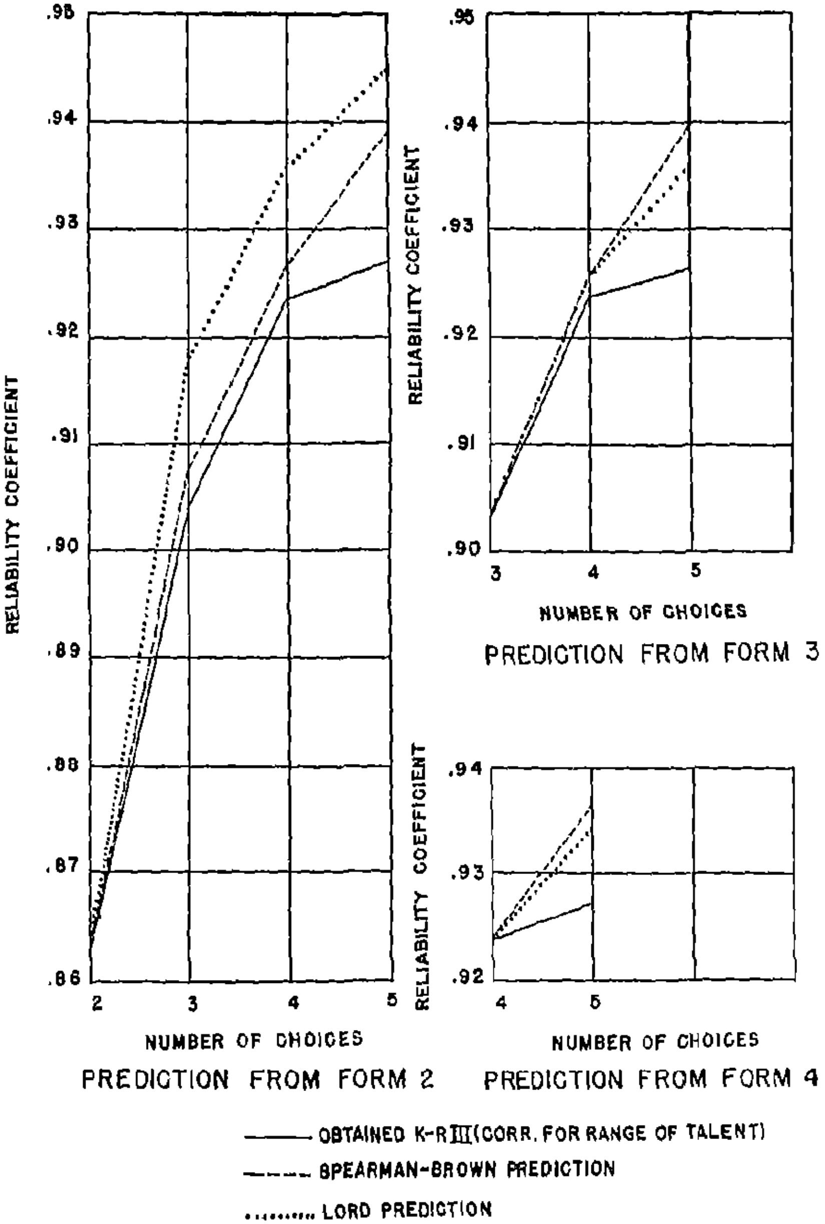
In predicting from Form 2, the Spearman-Brown is more accurate than Lord's formula. This is especially true of the predictions to Forms 3 and 4 from Form 2.

In predicting from Form 3 to Forms 4 and 5, however, the two formulas yielded the same estimate for Form 4 and Lord's formula was slightly more accurate for Form 5.

Finally, in predicting from Form 4 to Form 5, the Lord formula was slightly more accurate. Thus neither formula emerges with a clear-cut superiority.

We offer no rationalization for the failure of Lord's formula, especially derived for the use here made of it, to predict changes in reliability more closely than the Spearman-Brown formula. It is evident that Lord's assumptions are not met by our test. Nor would they be much more closely approximated by any mental ability test now in use. It is tempting therefore to ascribe its comparative failure to its unrealistic assumptions.

But the use of the Spearman-Brown formula is also vulnerable


PREDICTION FROM FORM 2

PREDICTION FROM FORM 3

PREDICTION FROM FORM 4

Figure 1. Obtained and Predicted Kuder-Richardson Case III Coefficients for Different Numbers of Choices.

since it was not derived for the use here made of it and since no rationale for its relative success has yet been offered. In increasing number of choices, we seem to do nothing strictly analogous to increasing number of items, number of minutes, or number of judges, to all of which the Spearman-Brown should and does apply. Yet the fact remains that when a two-choice test is changed to a three-choice or a four-choice test, we obtain reliability estimates very similar to those we would get if we had multiplied the number of items by 1.5 or by 2.

The practical significance of these findings for test builders is that test 'reliability' can be increased as predictably by increasing number of choices as by increasing number of items. If the author of a fifty-item two-choice test finds it has a reliability, by any of the formulas, of say .82, and wishes to increase this figure to, say, .90, he can choose between writing another fifty two-choice items, maximally equivalent to those already written, or writing an additional two valid choices for each already written item.

The difficulty and validity of the choices added would need to be carefully considered since they might add nothing if they were less plausible and valid than the choices already present in the items. But similar considerations would apply to whatever new items might be added. The decision between these paths to higher reliability might then depend on whether it was more feasible to add choices or to add items without changing the factorial composition of the test.

Loevinger's Homogeneity Formula.—To throw light on its behavior with a test of the kind for which it was designed (a power test of ability) we have computed Est H_t for our various forms.

The values of Est H_t (unadjusted for range of talent) in Table 4 show that this statistic increased in size as the number of choices increased, up to four choices, just as did the unadjusted K-R III reliability estimates. This suggests that, as the error of measurement of the individual items was decreased, a greater proportion of the total variance went into the general factor variance of the test.

But Est H_t did not increase regularly as the number of items increased from twenty-two to forty-five to ninety, for Form 5. Thus Est H_t increased from .265 to .311 for the change from twenty-two to forty-five items (Set A) but decreased from .311 to .307 for ninety items. K-R III on the other hand, increased regularly from twenty-two to ninety items. (It is noteworthy in passing that the changes in K-R III values from that for twenty-two items to those for forty-five and ninety items are in very close accordance with Spearman-Brown formula.)

TABLE 4.—VALUES OF LOEVINGER'S EST H_t AND K-R CASE III (NOT CORRECTED FOR RANGE OF TALENT) FOR TESTS DIFFERING IN NUMBER OF CHOICES PER ITEM, AND IN NUMBER OF ITEMS PER TEST

Type of Test	Loevinger's Est H_t	K-R Case III
45-item, 2-choice	.182	.863
45-item, 3-choice	.229	.898
45-item, 4-choice	.311	.928
45-item, 5-choice	.311	.928
22-item, 5-choice	.265	.843
45-item, 5-choice (A)	.311	.928
45-item, 5-choice (B)	.302	.926
90-item, 5-choice (A + B)	.307	.962

When the Kuder-Richardson coefficient and Loevinger's homogeneity index are computed on the other half (B) of the ninety-item test, the two estimates do increase together. This indicates that while Est H_t and K-R III need not vary together, it is possible for them to do so. Furthermore the numerical values of Est H_t are much lower than those of the 'reliability' coefficients for the same tests.

These results show that, although proposed as a 'partial alternative' to reliability, homogeneity does not reflect increased test length as reliability estimates do. This is perhaps understandable in that homogeneity depends on the persons' consistency in (1) failing all items requiring ability beyond a point of maximum difficulty and (2) passing all items below that point. If, as the number of items increases, the differences in difficulty between items become smaller, such consistent behavior becomes less probable, and the values of H would decrease. It follows that, of two tests differing only in their distributions of item difficulty, the one with the greater variability in item difficulty should yield the higher H.

We have looked into this possibility by computing H_t for two sets of nineteen items from our five-choice form. The distributions of difficulty (p) of the two sets differed as shown below:

Difficulty p	Number of Items Set A	 Set B
.60–.69	3	
.50–.59	3	
.40–.49	3	10
.30–.39	3	9
.20–.29	3	
.10–.19	3	
.01–.09	1	
Total items	19	19
No. of cases	225	225
Mean	7.38	7.04
S. D.	4.47	5.14
Est. H_t	.345	.310

The difference in H_t between Sets A and B, although apparently not great, is in the predicted direction. In the absence of sampling statistics for H_t, we have, of course, no adequate way of judging the significance of this difference.

It is evident that H_t although it has limits of zero and unity for perfect heterogeneity and homogeneity, respectively, does not have the same interpretability as do reliability coefficients. H_t was much lower than r_{tt} for a given test. How high H_t should be to indicate a 'high' degree of homogeneity we have as yet no way of knowing. What is needed is some kind of 'norms,' arrived at either empirically or through derivation of the 'standard error' of H_t, which will give us a basis for interpreting values of H_t.

SUMMARY

1) This research deals with two problems: (1) differences between single-trial estimates of reliability and homogeneity,

and (2) differences between two formulas for predicting change in reliability with changes in number of choices per test item.

2) The procedure was to administer four forms of Parts 1 and 2 of the O. S. U. Psychological Test, the forms differing in number of choices per test item, to four groups of high-school and college students. Values of the corrected odd-even, Guttman L_4, Kuder-Richardson Case III and Case IV, and Loevinger homogeneity statistics were computed from the resulting data. Adjustments were made for range of talent to make the reliability estimates comparable from form to form; controlled rather than random elimination of choices was used to make the various forms.

3) The differences between the obtained single-trial estimates of reliability were in the directions expected from analyses of the derivations of the formulas. More importantly, the differences here obtained with a fairly typical power test of ability were too small in size to be of practical significance.

4) The differences between the formulas (Lord and Spearman-Brown) for predicting change in reliability with number of choices did not clearly indicate the superiority of either. The accuracy of both compared favorably with that of the Spearman-Brown when applied to number of items.

5) The Loevinger homogeneity index increased as did the reliability estimates as number of choices was increased. Its numerical value was much smaller than that of the reliability coefficients. There was no basis for judging the psychometric or statistical significance of the values obtained. That is, did they indicate a relatively heterogeneous or homogeneous test?

REFERENCES

1) R. M. Adkins and H. H. Remmers. "Reliability of Multiple-Choice Measuring Instruments as a Function of the Spearman-Brown Prophesy Formula, VI." *Journal of Educational Psychology*, XXXII (1942), 385–390.

2) L. J. Cronbach. "Test 'Reliability': Its Meaning and Determination." *Psychometrika*, XII (1947), 1–16.

3) H. R. Denney and H. H. Remmers. "Reliability of Multiple-Choice Measuring Instruments as a Function of the Spearman-Brown Prophesy Formula, II." *Journal of Educational Psychology*, XXXI (1940), 699–704.

4) George A. Ferguson. *The Reliability of Mental Tests*. London: University of London Press, 1941, p. 176.

5) L. Guttman. "A Basis for Analyzing Test-Retest Reliability." *Psychometrika*, X (1945), 255–282.

6) ———. "A Basis for Scaling Qualitative Data." *American Sociological Review*, IX (1944), 139–150.

7) J. Milton House and H. H. Remmers. "Reliability of Multiple-Choice Measuring Instruments as a Function of the Spearman-Brown Prophecy Formula, IV." *Journal of Educational Psychology*, XXXII (1941), 372–376.

8) R. W. B. Jackson. "Reliability of Mental Tests." *British Journal of Psychology*, XXIX (1939), 267–287.

9) R. W. B. Jackson and G. A. Ferguson. *Studies on the Reliability of Tests.* Bulletin No. 12, Dept. Educational Research., University of Toronto, (1941).

10) T. L. Kelley. "The Reliability Coefficient." *Psychometrika*, VII (1942) 75–83.

11) G. F. Kuder and M. W. Richardson. "The Theory of the Estimation of Test Reliability." *Psychometrika*, II (1937), 151–160.

12) Jane Loevinger. "A Systematic Approach to the Construction and Evaluation of Tests of Ability." *Psychological Monographs*, LXI (1947) No. 4, pp. 49.

13) ———. "The Technic of Homogeneous Tests Compared with Some Aspects of 'Scale Analysis' and Factor Analysis." *Psychological Bulletin*, XLV (1948), 507–520.

14) Frederick N. Lord. "Reliability of Multiple-Choice Tests as a Function of Numbers of Choices per Item." *Journal of Educational Psychology*, XXXV (1944), 175–180.

15) T. M. Newcomb. "Social Attitudes and Their Measurement." Chapter 13 in Gardner Murphy, Lois B. Murphy, and T. M. Newcomb. *Experimental Social Psychology*, Revised Edition, New York: Harper & Bros., 1937.

16) H. H. Remmers and L. Whisler. "Test Reliability as a Function of Method of Computation." *Journal of Educational Psychology*, XXIX (1938), 81–92.

17) P. J. Rulon. "A Simplified Procedure for Determining the Reliability of a Test by Split-Halves." *Harvard Educational Review*, IX (1939), 99–103.

18) Herbert A. Toops. "The Validity and Difficulty of Items and Item Alternatives of Forms 18, 19, 20, 21 of the Ohio State University Psychological Test." Mimeographed. Dept. Psychol., Ohio State Univ., Columbus, Ohio.

19) D. A. Walker. "Answer Pattern and Score Scatter in Tests and Examinations." *British Journal of Psychology*, XXII (1931), 73–86.

20) ———. "Answer Pattern and Score Scatter in Tests and Examinations." *British Journal of Psychology*, XXII (1936), 301–308.

21) ———. "Answer Pattern and Score Scatter in Tests and Examinations." *British Journal of Psychology*, XXX (1940), 248–260.

GROUP GUIDANCE IN THE PROGRAM OF A READING LABORATORY

GLADYS L. PERSONS and MARY H. GRUMBLY
University of Bridgeport, Bridgeport, Conn.

The purpose of a reading laboratory or clinic is the rehabilitation of retarded students in skills of reading, study habits, self-confidence and personal development. Naturally, guidance should play a large part in this program. At the Laboratory of the University of Bridgeport, such a program has been initiated. The procedures and results will be discussed in this article.

Of the twelve principles and practices of Guidance as set forth by Cox, Duff and McNamara, it is apparent that six are fundamental to the philosophy of a reading laboratory and are in effect in varying degrees according to the size and duties of the staff. These six principles of guidance are:*

1) Guidance consists in helping pupils to set up objectives that are for them dynamic, reasonable, and worth while, and in helping them, so far as possible, to attain these objectives.

2) The major fields in which guidance is necessary are health, vocation, avocation, education, and human relations.

3) The idea of guidance is inherent in all efforts to educate.

4) The kind and amount of guidance needed varies greatly with different children and in different situations and at different times.

5) A research and measurement program is an essential part of successful guidance work.

6) The proper adaptation of curriculum and method to the needs of individual pupils is best promoted through guidance activities of teachers working in a democratically organized school system.

That these principles are in operation in the Reading Laboratory of the University of Bridgeport is evidenced by the importance placed on the case history in the diagnosis, the conferences with one or both parents, conferences with school principals and guidance directors and the eager acceptance of all school records and records of previous tests. It is apparent, too, in the testing which precedes instruction of each and every student. The following tests are administered (forms of some being adapted to age and achievement of individual tested):

---

* Cox, Duff and McNamara. *Basic Principles of Guidance*, New York: Prentice-Hall, Inc., 1948. Quoted by permission of the Publishers.

Test	Purpose of the Test
1) Stanford-Binet or Wechsler-Bellevue Test of Mental Ability	to determine intelligence quotient and to gain insight into personality
2) Draw-a-Person	to gain additional insight into personality
3) Iowa Silent Reading or Stanford Silent Reading or Gates Primary or Reading Readiness	to determine the silent reading level
4) Vocabulary—Inglis	to determine word knowledge
5) Standardized Spelling	to determine ability to spell and obvious causes of spelling difficulties
6) Telebinocular—a screening test of vision	to refer to an ophthalmologist if results are unsatisfactory
7) Gray Oral Check	to determine the level of oral reading and types of errors
8) Dominance	to determine if left-sidedness or mixed dominance is present
9) Ophthalmograph	to get a picture of eye movements during reading and of rate of reading
10) Achievement Test in common school subjects	to study growth and achievement of individual

Once the child has been accepted as a student, there is an attempt to see that his health and personality needs are met. There are interviews with doctors, eye and ear specialists and psychiatrists; contact is made with any and all who have been ministering to the child. Efforts are made to develop apparent talents and to nourish aggressive tendencies where only self-distrust and timidity exist. Instruction is planned to meet individual needs and re-testing at intervals to measure gains takes the place of subject-matter examinations and reports.

In order to meet individual needs in instruction, students are grouped according to their reading level and, whenever possible, according to their stage of physical and social growth. Because all students have some history of retardation before coming to the Reading Laboratory, there are no students below the age of eight and seldom any above the age of twenty in the morning or what is known as the full-time session. Student groups vary in number from one to eight or ten; they work with materials of interest to their age and of difficulty suitable to their reading level.

One typical group is that of five twelve-year old boys who have advanced in regular school as far as Grade IV but because of reading disability have little hope of going ahead. They are, therefore, in the Reading Laboratory. Two of them have better than average intelligence quotients; two are average. One is probably a mental defective whose personality makes up for much. (Those dealing with this student feel that once skills of reading are acquired, his placement in the intelligence test may be changed.) These boys are talking about vocational plans already and evince interest in occupations and their own capacity for profiting by training in certain vocational fields.

Every reading laboratory has a group of adolescent boys who were pushed through grade school and who then realized with a shock (as did their parents) that they were not prepared for high school even though they had been handed a grade-school diploma. Such sixteen-year-olds have constantly in mind the importance of preparing to earn a living. In addition to developing reading skills, they want to know what fields they may hope to enter, where they may obtain training for the vocation of their choice, and what the educational requirements are.

A typical evening group of students in the Reading Laboratory is composed of young working men who hope for advancement in their places of employment or who hope to take advanced courses which will lead them into other fields of work, both more interesting and more profitable. A realization that lack of reading skills is keeping them from their desired goals has brought them to the Reading Laboratory. Although through their daily experiences they develop knowledge of the working world, they want some help in broadening their understanding of fields of work and of those qualities which contribute to advancement.

The work of a reading laboratory, therefore, could be strengthened by placing more emphasis on educational and vocational guidance. Vocational interest could conceivably provide the motivating force in accelerating gains in reading skills, but, above all, knowledge of the working world is due these students because of the imminence of earning a living and assuming the duties of citizenship. They should be given some idea of their responsibilities as citizens and some idea of how our government works and of how much reading skill they need for their chosen work and for citizenship.

Teen-age boys and girls following re-training in reading, should have at least a tentative vocational goal, and should leave the reading laboratory for a school where they may expect training which will help them attain that goal. Therefore, some vocational aptitude testing should be planned for this group—not necessarily within the reading laboratory, itself.

A study of occupations, if introduced into the program of the reading laboratory, could contribute to the reading program, widen the horizons of these boys and girls as their reading skills are being developed, and fortify them with knowledge of themselves and the world of work which will help them and their parents to select the schools which they will enter following training in reading. Any such courses, however, must contribute, as well to the development of reading skills and must be planned to harmonize with the philosophy and method of the reading program.

The schools in the area of the reading laboratory should be known to the laboratory, particularly those schools where students may build on the reading skills developed in the reading laboratory. It is important to know, also, the schools where vocational training is given. Schools which develop special talents and yet do not require a high-school diploma, should be known to the laboratory. There should be knowledge, too, of educational advisement services and of the most readily available and reliable vocational advisement and testing service.

It appears that a group guidance course would be the logical means of helping the students of a reading laboratory develop their understanding of the world of work, develop some skill in the scrutiny of self and some ability to determine means and places for self-improvement. It should be the aim of such a course to bring about the increasing ability for self-guidance and self-direction which will stand by the student after he has left the relatively protective atmosphere of the reading laboratory.

In February of last year, such a course was introduced into the program of the Reading Laboratory at the University of Bridgeport. It was offered to a group of five adolescent boys, only one of whom had been a regular high-school student. These boys were at ease with one another, thus presenting no problems as to free discussion. They were thinking in terms of the future, and were, therefore, already motivated. The teacher had had training in the techniques of group guidance and had had experience in group work.

DATA ON A REPRESENTATIVE CLINIC GROUP

It was at the beginning of the new semester in February, that the course in group guidance was introduced. It was planned for two half-hour periods per week, coming the last period on Tuesday and Thursday. The group was programmed to read history with the same teacher at the same time on Monday and Wednesday, thus maintaining contact.

There follows a description of each student in the group:

I

Martin, who is sixteen years old, has been a tenth-grade student in a nearby high school where his grades have been D's with an occasional C. Both the Principal and the Guidance Director of the school and also his mother are interested in Martin, who is preparing for a business career. His mother has migraine headaches, and frequently withdraws from her children, insisting on quiet in the home. The father is frequently away from home, and when at home is unsympathetic toward the children whom he feels are not a credit to him.

Test results for Martin indicated that his intelligence quotient places him in the dull-normal range; his memory is reliable and even. He is immature emotionally and does little thinking for himself. He is high in acuteness of observation and ability to handle concrete problems; low in practical judgment. He has satisfactory vision at near point. His dominance is completely right. In oral reading he is slow and inaccurate and in silent reading he places on grade level 5.4—thus evidencing a retardation of five years. He reads simple material at 172 words per minute with good comprehension. The psychologist concluded that while Martin does not appear to have much capacity for educational tasks involving verbal problems, he appears to be able to do tasks involving design, mechanical drawing and other concrete skills.

II

Kevin, aged sixteen, is the son of a mother who is a college graduate and of a father who is a successful business executive. His sister is in college and is doing well. Kevin was very ill as an infant, but has had no serious illness since he was six months old. Kevin never learned to read. He repeated the first grade; has had his school situation changed frequently; has been tutored, and yet, he finally reached the eighth grade unable to read. He came to the Reading Laboratory in March, 1948, for training.

The tests administered indicate that his intelligence quotient is in the average range. He is emotionally immature, very unsure of himself and gives up easily. His test of vision indicated some difficulty at near point but the ophthalmologist does not believe that a correction is indicated. His dominance is chiefly right, although he does some things with his left hand. Oral reading of very simple material is accomplished in a painfully slow manner with much stammering; in silent reading he placed on grade level 3.2 and his graphs show reading of simple material at 128 words per minute with moderate comprehension.

Parental plans for Kevin are to have him enter high school for low-level students where he will take some commercial subjects and some trade training courses. Eventually, he will go into his father's business, which is the selling of brick.

III

Charles is fifteen and has been failing in all subjects in the eighth grade. He repeated the seventh grade. His trouble with reading began in the fourth grade. He has never been a behavior problem. His family find him a helpful member of the group but a reticent one. He has one brother, age seventeen, who is getting along fairly well in high school.

Charles' intelligence quotient is in the average range. He has a poor memory. Emotionally he is very immature and suffers greatly from feelings of inadequacy and insecurity. His test of vision shows trouble with lateral imbalance. His dominance is completely left, but he uses his right hand in golf. His oral reading is slow and punctuated with errors; his placement on a silent reading test is 5.7 and he reads 66 words per minute with good comprehension.

Charles' family would like to have him enter a vocational high school, but Charles has told his brother that he wants to go to a regular high school.

IV

Gilbert is sixteen. His mother is a college graduate; his father attended a university. His sister is in college at present. His brother, who is older, is a non-reader; he is successfully operating a farm. They have a good home life although there was much tension in the home during the depression years. Gilbert repeated the *first grade* and has since been pushed ahead until he reached the seventh grade without any significant accomplishment to his credit. However, he was always excellent in drawing.

Gilbert's intelligence quotient is in the dull normal range. He is a slow thinker and a slow learner. There are no indications of emotional disturbance. His vision is satisfactory at both near and far point. His dominance is completely right. In oral reading, he is very slow and inaccurate. On a silent reading test, he places at grade level 3.4 and he reads 100 words per minute with ninety per cent comprehension.

Gilbert is interested in carpentry and masonry and does such work around his home and on his brother's farm. He has good work habits; he is very reliable and systematic. He handles money well. His ability to draw is marked. The instructor from the Art Department of the University who has had this group of boys for sketching, says that Gilbert is very intelligent about following directions and that his achievement in this class is outstanding.

V

John is sixteen and the member of a very cohesive family. His sister is in college and his twin brother is in high school. John progressed through the grades of a private school without significant accomplishment and came to the Reading Laboratory in January, 1948, practically a non-reader.

Tests show his intelligence quotient in the average range. Marked emotional disturbance, insecurity and anxiety are present. He is aggressive, egocentric and somewhat hostile. His vision is satisfactory. In dominance, he is completely left. His oral reading is slow with many errors. At the time John became a member of the group guidance class, his silent reading was on a 5.2 grade level, his spelling on 4.7. He was reading 207 words per minute with excellent comprehension.

John has an aptitude for business and has done much work for his father, who is a building contractor. He understands money and the spirit of competition. He owns and drives a car. He is getting training in the various building trades through the work he is doing; the training is evidently planned. He can read blue prints to some extent, but both he and his family expect him to take some technical training in the drawing of plans. John has a *pleasant manner but is deficient in a sense of social responsibility.*

(———)

These five individuals are a typical reading laboratory group. They all need help in adjusting to the world of work as well as training in reading. Fundamentally, they have the same needs, but as individuals differ sufficiently so that they will benefit from the interplay of group thought and group discussion.

The major purposes of this group guidance course were determined as:

1) To give to the student certain basic concepts concerning responsibilities of citizenship, acceptable social behavior, success in any occupation.

2) To aid the student in obtaining adequate and accurate information about schools and courses so that he may have facts on which to base his decisions.

3) To present to the student a method of investigating occupations.

4) To assist the student in making tentative plans for the future.

To achieve the first purpose, lessons were planned using Unit Number 1 of the *Occupations Course* by R. Floyd Cromwell and Morgan D. Parmenter, published by Guidance Publishing Company, Limited, and distributed by the Psychological Corporation. Each student was provided with a text notebook and instructed in its use.

The instructor developed worksheets according to the methods of the Reading Laboratory which has a definite philosophy underlying its reading instruction. These worksheets were attached by each student to his text notebook, being placed following the chapter on which each was based.

The lessons based on the Cromwell and Parmenter text notebook (twenty-four in all) were carried on through group discussion, silent reading in class followed by tests of comprehension and oral reading. The problems selected for discussion and the cases analyzed by the group were chosen with the needs and interests of these particular boys in mind.

The lessons on occupations were developed using material published by the Bureau of Vocational Guidance, Connecticut Department of Education. Worksheets to be used with each of ten job analyses written for junior high-school students were prepared by the instructor. These worksheets were designed to accord with the philosophy and method of the Reading Laboratory and to point up the thinking of the student in regard to the important items to be considered when looking into any job. Each member of the group worked on his choice of three of the ten job analyses offered.

The last three lessons were designed to give the students information on schools in the Bridgeport area. Bulletins and other descriptive material were read by the students. Letters of inquiry were written by students to the special schools which seemed to offer training suited to their interests and aptitudes. The last lesson placed emphasis on the importance of tentative vocational plans, stressing the need for flexibility in these plans and the advantage of broad training over training in specific skills at this stage in their education.

Flashmeter work is part of the daily program of every student in our Reading Laboratory. It is used to quicken visual and mental perception and, therefore, material selected for reading is familiar to the class in subject matter and is in accord with their interests. The reading of a paragraph is, as a rule, the final part of the lesson. A series of ten paragraphs were prepared by the teacher conducting the group guidance course and presented by the teacher who had this group of students for flashmeter work. This served as review and as a means of emphasizing certain aspects of the course.

Good citizenship means voting, paying taxes, obeying the laws of the land; but more than that, it means a willingness to be informed and to study to some degree problems of government. To give these boys a concept of group relationships and of the responsibility of the individual to the group, together with a broad view of government functions, was necessary if the purpose of giving the student a basic concept of the responsibilities of citizenship was to be achieved.

Inasmuch as time did not permit this to be worked into the thirty lessons specifically allotted to group guidance, these lessons on citizenship were substituted for a number of history lessons. The text selected was *We Are The Government* by Mary Elting and published by Doubleday, Doran and Company, Inc. This is an accurate and interesting account of the various departments and branches of the Federal Government. It is written simply enough to be read by these boys, and yet it is sufficiently adult in format and presentation that they will not feel affronted. This is the text used for reading. Questions to test comprehension are prepared by the teacher.

The weekly assemblies of the student body at the Reading Laboratory provide occasions for these boys to present to the younger students their concepts of good citizenship and their broadening view of their country's development. This widens the horizons of the younger students, motivates the members of the group guidance class and gives them practice in expressing their opinions in public.

One device used in the Reading Laboratory to encourage independent reading, is a large file card on which the student keeps a record of the books he has read. A shelf of books from the library on subjects related to this course was placed in the group guidance classroom. Students were encouraged to select books for independent reading from this shelf and to add it to their record of books read. They were invited to summarize and evaluate these books for other members of the group.

Most group guidance courses plan some aptitude testing for the members of the course. Since this was in the tradition of such courses, three paper-and-pencil tests were administered by the instructor to this experimental group in the Reading Laboratory. The tests administered were the Revised Minnesota Paper Form Board Test, the Bennett Test of Mechanical Comprehension, and the O'Rourke Mechanical Aptitude Test, Junior Grade. Nothing definitive was indicated by the results of these tests. Arrangements have now been made to administer Performance Tests of Aptitude to these students through the Personnel Division of the University of Bridgeport.

CONCLUSIONS

Work with this group of adolescent boys during these three months suggests that a group guidance course should logically form a part of the program of the Reading Laboratory.

1) Interest in learning about occupations and the qualities which contribute to success in the world of work, motivates reading and, therefore, contributes to the development of reading skills in the student.

2) The material used is on an adult level and, yet, within the ability of these students to read under guidance; therefore, it fills a need constantly felt in the Reading Laboratory—that of providing material which takes into account the interests of students, which is presented in a manner and format suited to their age, and which is within the range of their reading ability.

3) The work of this course stimulates thinking along adult lines and, thus, contributes to the maturing process of these boys, a need usually revealed in the initial testing.

4) The lessons on citizenship fill a definite need in that they contribute to the maturing process of the student, give him background material which he missed in his regular schooling because of deficient reading, and stimulates his thinking and reading along broader lines.

5) Performance tests rather than paper-and-pencil tests should be given to determine the individual aptitudes of the

members of such a group. This should be done by experts who are working regularly in this field.

6) Through this group guidance course the student is helped to choose a tentative goal and to make plans for training upon finishing his work at the Reading Laboratory, thus making it more likely that the skills and work habits acquired in the Laboratory will be maintained and developed.

RELIABILITY AND VALIDITY OF INVOLUNTARY BLINKING AS A MEASURE OF EASE OF SEEING*

MILES A. TINKER
University of Minnesota

In any measuring device, the problems of reliability and validity are of high importance. The reliability of any measuring instrument is the consistency with which it measures the kind of response or behavior involved. Validity of a technique depends upon the fidelity with which it measures whatever it purports to measure. Only when adequate reliability and validity of an instrument, test or technique have been demonstrated, can one place confidence in results obtained by that measuring device.

Rate of involuntary blinking has been extensively used by Luckiesh[10] and Luckiesh and Moss[14] as a criterion of ease of seeing and of readability of print. In fact data obtained by this technique (rate of blinking) form much of the bases for conclusions and recommendations concerning adequate illumination intensities and desired typographical arrangements. Luckiesh, Guth and Eastman[12] point out that researches by Luckiesh and by his co-workers employing the blink technique over a period of years have yielded useful data in the study of level of illumination and of typography. In another place Luckiesh[9] states that rate of involuntary blinking while reading is a very sensitive criterion of ease of seeing. He also claims in this report that rate of blinking while performing the same task under different conditions (as changes in level of illumination) is by far the most sensitive, significant and practical criterion of ease of seeing that has been used.

Since rate of blinking is considered such an important criterion of ease of seeing, it is of considerable importance to evaluate the reliability and validity of the technique. It might be noted further that, in using this technique, the certainty of discovered differences should be evaluated statistically. The purpose of the

---

* The writer is grateful to the University of Minnesota Graduate School for a research grant to finance this study.

present study is to present and coördinate the results obtained in a number of my studies on reliability and validity of the blink technique as a measure of ease of seeing.

RELIABILITY

In an initial study[21] the blink rates were recorded for each successive five-minute period during thirty minutes of reading for two groups: seventy-four adult readers and sixty-four adult readers. Conditions of the experiment were comparable to those specified by Luckiesh, Guth and Eastman.[12] Blink rates in both sets of data tended to increase slightly in the successive five-minute periods of reading. The blink-rate consistency or reliability coefficient was fairly high from one five-minute period to the next period, i.e., mean about .87. The trends were about the same for adjacent ten-minute periods of reading (.85). From initial to final five minutes (of the thirty minutes), the reliability dropped to about .50. For the ten-minute periods of reading at the beginning and end of the thirty-minute period, the reliability was about .65. For the first 15 minutes vs. the last 15 minutes the reliability coefficient was .75.

In a second investigation,[22] frequency of blinking was recorded while reading lower-case text and all capital text for ten minutes each during each of two experimental sessions. There were sixty subjects. Records were kept for each five minutes of the ten-minute periods. On comparing the reliability of the blink frequency from one session to the next, the coefficients were uniformly high: .86 for five-minute periods; .89 for ten-minute periods; and .94 for twenty-minute periods.

A third study[24] was concerned with blink rate in reading book type and newsprint. There were sixty subjects. Records were kept for each five minutes of the ten-minute reading periods. For the five-minute periods the reliabilities were .82 and .84; for ten-minutes it was .95.

In a final study,[23] forty-two subjects read for fifty-five minutes under two footcandles of light and for a like period under one hundred footcandles. Blink frequency was recorded for the initial and final five minutes of each period. For reading under two footcandles the reliability (first vs. final five minutes) was .87; for one hundred footcandles, .57.

When data in all four studies are considered, trends in reliability coefficients are as follows: For successive five-minute or ten-minute periods of reading, and for five-minute or ten-minute periods of reading at two separate experimental sessions, the coefficients are relatively high—around .85 to .95. This represents highly satisfactory consistency of performance. When initial and final five minutes of thirty to fifty-five minutes of reading are compared, the reliabilities fluctuate around .50. But in one case out of four a coefficient similarly derived was .87. Ten minutes of reading yields slightly higher reliability than five minutes. It may be concluded that consistency of performance in measuring blink frequency during reading is adequate where group comparisons are involved although these coefficients at times barely reach the minimum requirements when initial and final five minutes of a lengthy period of reading are compared. For any short period of five to fifteen minutes of reading, the consistency or reliability of blink frequency is highly satisfactory.

In most other studies of blink frequency, reliability of performance has not been cited. In Carmichael and Dearborn's investigation,[5] groups of ten subjects each read on different occasions various kinds of material for six hours. Blinks were recorded for the first five minutes of reading and for the last five minutes of the successive twelve half-hour intervals. Correlations were computed between initial scores and the scores at the end of each successive half hour. These (reliability) correlations reveal how consistently the subjects maintained their original positions in the group. Although these coefficients revealed wide variability, ranging from .06 to .99, four-fifths of them were .70 or above. Very few were below .50. In most of their samples, therefore, the reliability of the blink technique seems adequate.

VALIDITY

In various reports,[9,10,11,12,13,14] Luckiesh and his coworkers have employed and supported rate of involuntary blinking as an adequate measure of readability of print, of ease of seeing, visual efficiency, ocular fatigue, etc. The assumption is that rate of involuntary blinking increases with the severity or difficulty of the task and thus reflects the degree of effort expended by the subject in performing the visual task. The validity of a measure should be established in terms of how well it agrees with a criterion. And a criterion is ordinarily considered to be some measure which experience has shown to be an adequate or true measure of the trait under consideration.

Three studies have been completed in our laboratory to check the validity of blink rate as a measure of ease of seeing. In the first investigation,[22] the relative readability of text in all-capital printing was compared with text in lower-case. Earlier studies have established that all-capital text is significantly less readable than text in lower-case. Tinker and Paterson[26] demonstrated that all-capital text is read about twelve per cent slower than lower-case text. The same trend was discovered in an eye-movement study.[27] Measuring the trend in blink rate in a similar comparison will, therefore, furnish a check on blink rate as a measure of readability.

The experiment was carried out in a light laboratory with ten footcandles of well diffused illumination. Sixty university students served as readers for two experimental sessions, reading lower-case text followed by all-capital text at one session and in the reverse order at the second session. A counter-balanced experimental design was used and the subjects were adapted to the illumination prior to the reading. Comparisons were made for five, ten and twenty minutes of reading. Both number of blinks per period and rate of reading were measured. The findings *revealed no significant differences in rate of blinking while* reading text in all-capitals in comparison with lower case. But in all comparisons, the text in all-capitals was read significantly slower than text in lower-case, i.e., 9.53 to 19.01 per cent slower. The findings suggest that frequency of blinking is an unsatisfactory criterion of readability of print.

A second study[24] was concerned with the readability of book print and newsprint in terms of blink rate. In earlier work, Paterson and Tinker[16] have shown that newsprint is read significantly slower than book print. All evidence[14, 17] indicates that ease of seeing large (book) type is greater than for small (newsprint) type. A comparison of blink rates for these materials, therefore, should provide a check on the validity of blink rate as a measure of ease of seeing. In the present study the book type was twelve-point; the newsprint, seven-point. Specifications *for experimental design and procedures laid down by* Luckiesh were followed. For details see the original report.[24] The results show significantly fewer blinks (61.27 vs. 69.37) for reading the newsprint. This trend is directly opposite to what should have occurred if rate of blinking is an adequate measure of ease of seeing, i.e., there should have been more blinks in reading the newsprint. Obviously, further checking of the validity of the blink technique is needed.

In the third study[25] concerned with influence of illumination intensity, the specifications of Luckiesh, Guth and Eastman[12] and Luckiesh and Moss[13] concerning experimental design and procedures were followed in detail and the same reading material was used. Forty-two university students read for fifty-five minutes under two and under one hundred footcandles of light. Frequency of blinking during the first five and the last five minutes of the fifty-five-minute period were recorded. Details of conditions and procedures are given in the original report. The problem in this experiment is to check whether frequency of blinking increases significantly more for reading fifty-five minutes under the dim than under the bright light.

TABLE I.—CHANGES IN RATE OF BLINKING WHEN READING FIFTY-FIVE MINUTES UNDER TWO AND UNDER ONE HUNDRED FOOTCANDLES OF LIGHT. N = 42 READERS.

Foot-Candles	First 5 Min.		Last 5 Min.		r	Diff.	% Diff.	$\frac{D}{\sigma \text{Diff.}}$
	Mean	SD	Mean	SD				
2	25.8	23.2	33.9	27.9	.87	8.1	31.6	5.60
100	26.8	19.4	35.4	24.5	.57	8.6	32.0	3.97

The basic data are given in Table I. Examination of the means reveals that there is an appreciable increase in number of blinks from initial to final five minutes of reading under both levels of illumination. These changes in blink rate are significant at the one per cent level as shown by the critical ratios in the last column of the table. The same level of significance is obtained when the *t*-test is applied to the data.

Under two footcandles, the increase in blink rate is 8.1 or 31.6 per cent. Similarly, under one hundred footcandles, the increase is 8.6 or 32.0 per cent. There is, therefore, an approximately identical amount of increase in blink rate for reading about an hour under two and under one hundred footcandles.

It is universally recognized, in terms of literature on the subject, that reading under two footcandles is a more severe visual task than reading under one hundred footcandles. There is no trend in the blink data of this experiment to support this view. Under the conditions of this experiment, therefore, rate of involuntary blinking is not a valid criterion of ease of seeing in reading.

DISCUSSION

Reliability or consistency of performance in rate of involuntary blinking appears adequate. In those studies where the reliability has been computed, the coefficients tend to run fairly high for most samples. Nevertheless, since low reliability occurs occasionally, the reliability should be computed in each investigation where the blink technique is used.

Validity of the blink technique as a measure of ease of seeing, however, is in question. In Luckiesh and Moss[13] and Luckiesh,[10] the findings reported uniformly show increases in rate of blinking under conditions which presumably involve more difficult visual tasks such as with reduction in size of type, intensity of illumination, uniformity of lighting, etc. They conclude that blink rate as a criterion of ease of seeing is uniquely satisfactory. But check studies have failed to obtain data confirming the trends published by these authors.

McFarland, Holway and Hurvich,[15] in view of their findings and analyses, conclude that a high blink rate need mean neither an increase in visual fatigue nor an increase in difficulty of seeing. In his study, Bitterman[1] found a decrease in work output for reading six-point in comparison with ten-point type but no significant difference in blink rate for reading the two sizes of type. In the same study he found no significant difference in blink rate for reading under three and under ninety-one footcandles of light. In both comparisons the slight but insignificant differences were in favor of the more severe visual task which is contrary to the trends found by Luckiesh and his co-workers. Bitterman concludes that rate of blinking cannot be accepted as a criterion of readability and ease of seeing. In a later analysis Bitterman and Soloway[4] demonstrated that expenditure of greater effort in mental work is not necessarily reflected in an increased frequency of blinking.

Simonson and Brozek,[10] in studying the effects of illumination level on visual performance and fatigue, employed the severe task of visual discrimination in perceiving very small letters under two, five, fifteen, fifty, one hundred and three hundred foot-candles of illumination. Work output increased significantly in going from two, five or fifteen footcandles to the higher levels of illumination but there were no significant changes in blink rate. They conclude that the blink technique cannot be employed as a fatigue index. In a later study on the effect of spectral quality of light on visual performance and fatigue, these same authors[20] found that blinking rate did not show significant changes between illuminants although there were significant differences in certain performance and fatigue tests. In Carmichael and Dearborn's experiment[5] samples of blinking were taken at half-hour intervals during six hours of normal reading and microfilm reading. There were more blinks for normal reading than for the microfilm reading although most of the differences were not significant. Conclusions are not clear here for adults read microfilm faster and high-school subjects were more rapid in the normal reading. In studying the influence of typographical variations on reading by visual handicapped children, McNally[16] found no significant differences in frequency of blinking from one kind of text to another. Rate of reading revealed no significant differences either. Hence the whole set of data seems equivocal.

When all studies on the blink technique as a measure of ease of seeing are examined the following trends are revealed: (A) Luckiesh and his co-workers always report data that appear to indicate a higher rate of blinking for situations where the visual task is more exacting. (B) In all other investigations either (a) there are no significant differences in frequency of blinking in going from easy to severe visual tasks or (b) the trends of the data are equivocal. (C) There are, therefore, serious doubts concerning the validity of the blink technique as a measure of ease of seeing.

In reply to criticisms[1, 22] of the blink technique, Luckiesh[11, 12] has argued that workers in other laboratories have not duplicated his experimental set-up and methods of procedure, nor maintained the same reading attitudes in the subjects. Nevertheless, inability to confirm Luckiesh's results cannot be due to these factors for Tinker's recent experiment[25] duplicated all conditions and arrangements specified by Luckiesh. The data revealed no significant differences in blink rate for reading under two vs. one hundred footcandles of light.

Luckiesh[11] and Luckiesh, Guth and Eastman[12] imply that the increase in frequency of blinking from the beginning to the end of a lengthy period of reading supports the view that the technique is valid as a measure of ease of seeing. They point out that when conditions are comparable to those in their experiments, findings are similar. The following data were cited (per cent increase in blink rate is given in parentheses at end of items):

Luckiesh and Moss (N = 11), 60 min. of reading....... (31)
Hoffman (N = 30), 60 min. of reading................. (27)
Tinker (N = 74), 30 min. of reading.................. (25.2)
Tinker (N = 64), 30 min. of reading.................. (36.2)

It is obvious that the trends toward increase in frequency of blinking from beginning to end of a period of reading agree. It is doubtful, however, that this trend implies anything concerning the validity of the blink technique as a measure of ease of seeing. This trend toward an increase in blink rate is found in practically every experiment. For instance, Tinker[16] found an increase of 31.6 and 32 per cent while reading fifty-five minutes under two and one hundred footcandles, respectively. And Carmichael and Dearborn[5] found increases of about twenty-one, twenty, thirty-nine, and thirteen per cent for reading from two to six hours under sixteen footcandles. Many of these differences are statistically significant. Hoffman[7,8] has suggested that the change (presumably of a deleterious nature) in blink rate with continuous reading is the effect of a work factor (work decrement phenomenon). This seems a reasonable inference. Apparently Carpenter,[6] in his study of changes in rate of blinking during visual search, has interpreted this work decrement phenomenon as validating the technique as a criterion of ease of seeing. His increase in blinks from the first to the last half hour of a two-hour period was forty-three per cent. This is somewhat greater than the increases (about thirty-two per cent) found for one to two hours of reading.

In any case, if differential increase in frequency of blinking due to visual work is to be employed as a criterion of ease of seeing, the increase during a severe visual task should be greater than during an easy task by a statistically significant amount. No independent worker, however, has been able to confirm the results reported by Luckiesh and his co-workers when such comparisons have been made.

A word may be added concerning the statistical treatment of blink-rate data. Tinker[23] has questioned the appropriateness of the geometric mean as employed with these data by Luckiesh and his co-workers. The same criticism is offered by Hoffman.[8] He also severely criticizes the use of the percentage technique employed by them, and for basing conclusions on percentage differences rather than on raw score differences. Percentage differences are notoriously unreliable. Furthermore, if the raw scores are below 100 (as most of them are), percentages magnify the differences. Thus insignificant raw score differences may seem large when put into percentage differences. Furthermore, all obtained differences should be evaluated for statistical significance.

The trend of the data reported here plus a survey of related experiments raises serious questions concerning the validity of the blink technique as a measure of ease of seeing and readability of print. On the one hand are the data reported by Luckiesh and his co-workers. They[12] conclude that frequency of blinking "while reading is a very sensitive criterion of ease of seeing and that the results are significantly related to tenseness and other results of reading and other critical seeing." On the other hand, no independent worker, even when duplicating the experimental design of Luckiesh, has been able to confirm the results of Luckiesh and his co-workers. These independent workers all find that increasing the severity of the visual task has no differential effect upon frequency of blinking. Although the frequency of blinking does increase as reading is continued for thirty minutes to an hour or more, the rate of increase is no greater for severe than for easy visual tasks. Lack of confirmation of the Luckiesh findings can no longer be assigned to differences in experimental design.[25] The basis for the apparent consistency in the data from Luckiesh's laboratory and which cannot be confirmed by others is not clear at this time.

In any case, failure to substantiate the rate of blinking as a criterion of ease of seeing and of readability must lead to its rejection as an adequate technique in studies of visual work. Bitterman and Soloway[3] tend to agree with these conclusions. This would mean that much of the evidence put forward by Luckiesh[10] and Luckiesh and Moss[14] to support recommendations for desirable practice in the field of illumination and typographical arrangements is invalid and that the recommendations, therefore, are without sound foundation.

Various techniques, including frequency of blinking, have been promoted as measures of efficiency in terms of the physiological cost of the work. None of these have been satisfactory. It is generally admitted, however, that an adequate measure of energy expenditure during visual work would yield important data to supplement work-output results. In the absence of the former it is feasible[2] to rely upon data from performance measures until more sensitive and valid indices of cost have been developed.

SUMMARY AND CONCLUSIONS

1) The results of four experiments concerned with reliability and validity of frequency of blinking as a measure of ease of seeing are reported.

2) Consistency of response or reliability of blink rate is adequate as used in experiments on readability and ease of seeing.

3) Frequency of blinking is not a valid measure of readability and ease of seeing.

REFERENCES

1) M. E. Bitterman. "Heart rate and frequency of blinking as indices of visual efficiency." *J. exp. Psychol.*, 1945, 35, 279–292.

2) ———. "Transfer of decrement in ocular tasks." *Amer. J. Psychol.*, 1946, 59, 422–438.

3) ——— and E. Soloway. "Frequency of blinking as a measure of visual efficiency: some methodological considerations." *Amer. J. Psychol.*, 1946, 59, 676–681.

4) ———. "The relation between frequency of blinking and effort expended in mental work." *J. exp. Psychol.*, 1946, 36, 134–136.

5) L. Carmichael and W. F. Dearborn. *Reading and Visual Fatigue.* Boston: Houghton Mifflin, 1947.

6) A. Carpenter. "The rate of blinking during prolonged visual search." *J. exp. Psychol.*, 1948, 38, 587–591.

7) A. C. Hoffman. "Eye movements during prolonged reading." *J. exp. Psychol.*, 1946, 36, 95–118.

8) ———. "Luckiesh and Moss on reading illumination." *J. appl. Psychol.*, 1947, 31, 44–53.

9) M. Luckiesh. "On the rate of involuntary blinking." *Illum. Engng.*, 1944, 39, 60–71.

10) ———. *Light, vision and seeing.* New York: Van Nostrand, 1944.

11) ———. "Reading and the rate of blinking." *J. exp. Psychol.*, 1947, 37, 266–268.

12) ———, S. K. Guth, and A. A. Eastman. "The blink-rate and ease of seeing." *Illum. Engng.*, 1947, 42, 584–588.

13) ——— and F. K. Moss. "A summary of researches involving blink-rate as a criterion of ease of seeing." *Illum. Engng.*, 1940, 35, 19–32.

14) ———. *Reading as a visual task.* New York: Van Nostrand, 1942.

15) R. A. McFarland, A. H. Holway, and L. M. Hurvich. *Studies of visual fatigue.* Boston: Graduate School of Business Administration, Harvard Univ., 1942.

16) H. J. McNally. *The readability of certain type sizes and forms in sight-saving classes.* Teachers College Contrib. to Educ., No. 883. New York: Bureau of Publications, Teachers College, Columbia Univ., 1943.

17) D. G. Paterson, and M. A. Tinker. *How to make type readable.* New York: Harper and Bros., 1940 (available from the authors).

18) ———. "The relative readability of newsprint and book print." *J. appl. Psychol.*, 1946, 30, 454–459.

19) E. Simonson and J. Brozek. "Effects of illumination level on visual performance and fatigue." *J. opt. Soc. Amer.*, 1948, 38, 384–397.

20) ———. "The effect of spectral quality of light on visual performance and fatigue." *J. opt. Soc. Amer.*, 1948, 38, 830–840.

21) M. A. Tinker. "Reliability of blinking frequency employed as a measure of readability." *J. exp. Psychol.*, 1945, 35, 418–424.

22) ———. "Validity of frequency of blinking as a criterion of readability." *J. exp. Psychol.*, 1946, 36, 453–460.

23) ———. "Illumination standards for effective and easy seeing." *Psychol. Bull.*, 1947, 44, 435–450.

24) ———. "Readability of book print and newsprint in terms of blink-rate." *J. educ. Psychol.*, 1948, 39, 35–39.

25) ———. "Involuntary blink rate and illumination intensity in visual work." *J. exp. Psychol.*, 1949, 39, 558–560.

26) ——— and D. G. Paterson. "Influence of type form on speed of reading." *J. appl. Psychol.*, 1928, 12, 359–368.

27) ——— and D. G. Paterson. "Influence of type form on eye movements." *J. exp. Psychol.*, 1939, 25, 528–531.

THE ORIGIN AND DEVELOPMENT OF THE SPANISH ATTITUDE TOWARD THE ANGLO AND THE ANGLO ATTITUDE TOWARD THE SPANISH*

GRANVILLE B. JOHNSON, JR.
Arizona State College at Flagstaff

The problems derived from intergroup attitudes are varied and their solution is particularly important at the present time from the standpoint of national and international maladjustment. Research which probes the origins and traces the development of racial attitudes at the molar level would make a positive contribution to developmental and social psychology.

Underlying the present investigation is the explicit assumption that any measurable differences of personality between races or ethnic groups are due to an environmental and not a hereditary function. We are in agreement with the conclusions of Boaz:

"The occurrence of hereditary mental traits that belong to a particular race has never been proven. The available evidence makes it much more likely that the same mental traits appear in varying distribution among the principal racial groups. The behavior of an individual is therefore determined not by his racial affiliation, but by the character of his ancestry and his cultural environment. We may judge the mental characteristics of families and individuals, but not of race."[1]

THE PROBLEM

The purpose of the present investigation was to ascertain the degree of prejudice in progressively mature age groups of Spanish and Anglo subjects utilizing the Projective Test of Racial Attitudes and to reveal something of the origin and development of racial attitudes.

---

* This is the second in a series of three articles concerned with the origin and development of racial attitude. The previous paper, "An Experimental Projective Technique for the Analysis of Racial Attitude," appeared in April issue of this JOURNAL (41:4, p. xxx). In this early article the properties of the Projective Test of Racial Attitude, the major tool in the present investigation, were analyzed.

MATERIAL AND SUBJECTS

Materials.—1) The Projective Test of Racial Attitudes composed of six carefully selected pictures relevant to race situations and provocative of projected response was utilized for the analysis of racial attitudes. There was a hero figure with whom the subject identified himself and thus projected something of the dynamics of his personality into the depicted situations.

The situations conducive to projected response were identical except for the ethnic group variable. For Anglo subjects there were six cards of total Anglo content and six with an Anglo hero in a Spanish situation. For Spanish subjects the situations were reversed with a Spanish hero in an Anglo situation in six cards with identical cards without the Anglo element.

2) A tabulation sheet was utilized for recording the dynamics of responses.

Subjects.—1) Anglo: thirty male subjects (four-, eight- and twelve-year levels, N = 90) from public schools in a southwestern town.

2) Spanish: thirty bilingual male subjects (four-, eight- and twelve-year levels, N = 90) from public schools in a southwestern town.

A further limitation was that members of both groups were required to have lived in this specific town for the year prior to testing.

DEVELOPMENTAL CURVES

The first phase of the analysis was to trace the development of the Spanish attitude toward the Anglo and the Anglo attitude toward the Spanish. This was accomplished by computing the total number of responses in the Spanish-Spanish (Anglo-Anglo) columns and the Spanish-Anglo columns at each age level, then finding the significance of the difference between them. The change in the relative number of responses and resultant change in the significance of the difference was noted with increase in age.

Spanish.—Table I presents the significance of the difference between same group and not same group of ninety Spanish subjects, four, eight, and twelve years of age on the twelve cards of the Projective Test of Racial Attitudes. The difference indicated greater reaction in the Spanish-Anglo situations.

The analysis included: Effect of Environment (Frustrating); Reaction to Environment, (Frustration); Adequacy of Principle Character, (Superior); Ending Hero, (Defeat); and Ending Theme, (Unsatisfactory).

TABLE I.—SIGNIFICANCE OF DIFFERENCE BETWEEN SPANISH-SPANISH AND SPANISH-ANGLO RESPONSES USING THIRTY SPANISH SUBJECTS AT THE FOUR-, EIGHT- AND TWELVE-YEAR LEVELS IN THE PROJECTIVE TEST OF RACIAL ATTITUDES*

Age Levels	Eff. of Env. Frustrating	Reac. to Env. Frustration	Adeq. of Pr. Char. Superior	Ending Hero Defeat	Ending Theme Unsatisfactory
12	4.05	5.57	5.00	2.49	4.86
8	3.97	5.33	4.38	2.54	4.42
4	.58	.68	.81	2.74	.91

* Difference indicates greater reaction in Spanish-Anglo situations.

With reference to Table I it is quite evident that what happens in the temporal interval between the fourth and eighth years *contributes much more to development of racial attitudes than* the subsequent years studied. It is observed, too, that the organism's tendency to react to the environment in a manner betraying frustration manifests the greatest change in the first four years. It is important to note that this category is one of the most highly related with the total prejudice score and therefore is one of the most indicative of prejudice.

In both the Spanish and Anglo groups it is seen that the Spanish hero is defeated in the Spanish-Anglo situations as frequently at the four-year level as at the two older age levels. At all age levels the hero or character with whom the subject identified himself is defeated more often in the not same group than he is in the same group. It is possible to present the hypothetical explanation with reference to this finding that, though the result is the same, the cause may arise from different sources. With reference to this finding, defeat of the principle character by a member of the other ethnic group may be considered a criterion of prejudice.

Table II presents the significance of the difference between same group and not same group of total Spanish subjects at the chosen age levels. The analysis includes: Effect of Environment, (Neutral); Adequacy of Principal Character (Equal); Ending Hero, (Indeterminate); Ending Theme, (Indeterminate). The difference indicates greater reaction in the Spanish-Spanish situations.

TABLE II.—SIGNIFICANCE OF DIFFERENCE BETWEEN SPANISH-SPANISH AND SPANISH-ANGLO RESPONSES USING THIRTY SPANISH SUBJECTS IN EACH AGE GROUP NEUTRALITY CATEGORIES*

Age Levels	Eff. of Env. Neutral	Reac. to Env. Equal	Ending Hero Indeterminate	Ending Theme Indeterminate
12	−3.21	9.61	7.79	1.51
8	3.31	9.03	7.12	1.18
4	.23	2.19	1.52	0

* Difference indicates greater reaction in Spanish-Spanish situations.

A neutral, equal, or indeterminate response displays absence of affect to the specific situation. Though these particular categories do not reveal the direction of prejudice swing, they do demonstrate positive or negative reaction to the depicted situation. In Table II it is noted that the shift from neutrality in the Spanish-Spanish categories to positive or negative reaction in the Spanish-Anglo categories has a marked relationship with age.

Reflecting the organism's tendency to be affected by the Spanish-Anglo environment in either a positive or negative manner, the neutral column shows the greatest difference between Spanish-Spanish and Spanish-Anglo categories. This factor would possibly indicate that discrimination of and sensitivity to different aspects of the environment are prerequisite to more complex prejudice manifestations. It has been found that there is a positive relationship between prejudice and ability to recognize stereotypes.[1]

In 'Ending Theme,' (Spanish-Spanish), because no indeterminate responses were elicited at the four-year level (though there were six at the Spanish-Anglo level) no critical ratio could be ascertained.

The critical ratios in Table III represent the significance of the difference between Spanish-Spanish and Spanish-Anglo responses for the Spanish age levels. The difference indicates greater reaction in the Spanish-Spanish categories in the 'Effect of Environment,' 'Reaction to Environment,' and 'Ending Theme.' It indicates greater reaction in the Spanish-Anglo categories in the 'Adequacy of the Principle Character,' and 'Ending Hero.' The analysis included: Effect of Environment, (Helpful), Reaction to Environment, (Satisfactory); Adequacy of Principle Character, (Subordinate); Ending Hero, (Victory); Ending Theme, (Satisfactory).

TABLE III.—SIGNIFICANCE OF DIFFERENCE BETWEEN SPANISH-SPANISH AND SPANISH-ANGLO RESPONSES USING THIRTY SPANISH SUBJECTS IN EACH AGE GROUP IN THE PROJECTIVE TEST OF RACIAL ATTITUDES

Age Levels	Eff. of Env. Helpful*	Reac. to Env. Satisfactory*	Adeq. of Pr. Char. Subordinate†	Ending Hero Victory†	Ending Theme Satisfactory*
12	−1.89	−6.06	3.75	5.56	−4.82
8	−1.47	−5.33	3.59	4.25	−4.65
4	− .71	− .68	2.58	.16	−1.67

* Difference indicates greater reaction in Spanish-Spanish situations.
† Difference indicates greater reaction in Spanish-Anglo situations.

With reference to the data in Table III, it is noted that only Ending Hero, (Victory), and Adequacy of Principle Character (Subordinate), have more positive reaction in the Spanish-Anglo situation than in the Spanish-Spanish. This means, then, that both ascendance and submission become more manifest with age and may be considered criteria of prejudice. Though subordination of the hero in the Spanish-Anglo situation does develop with age, the difference is already statistically significant at the four-year level. Evidently, this feature is one of the first to develop with ego evolvement and group identification.

As age increases, it is noted, the reaction to the Spanish-Anglo environment becomes less satisfactory as does the feeling that elements of this environment are helpful. The ending of the theme also ends less optimistically.

Anglo.—The previous data demonstrated the phases and characteristics of attitudinal development of the Spanish subjects. The following analysis illustrates how and in what different age levels the Anglo attitude develops in adjustment to the Spanish.

Table IV presents the significance of the difference existing between Anglo-Anglo reaction and Spanish-Anglo reaction of the total Anglo group. This table includes: Effect of the Environment, (Frustrating); Reaction to Environment, (Frustration); Adequacy of Principle Character, (Superior); Ending Hero, (Defeat); and Ending Theme, (Unsatisfactory).

TABLE IV.—SIGNIFICANCE OF DIFFERENCE BETWEEN ANGLO-ANGLO AND SPANISH-ANGLO RESPONSES USING THIRTY ANGLO SUBJECTS IN EACH AGE GROUP ON THE PROJECTIVE TEST OF RACIAL ATTITUDES*

Age Levels	Eff. of Env. Frustrating	Reac. to Env. Frustration	Adeq. of Pr. Char. Superior	Ending Hero Defeat	Ending Theme Unsatisfactory
12	5.27	6.16	7.95	2.32	6.34
8	1.85	3.05	3.37	2.10	5.63
4	1.84	2.07	2.03	1.58	4.78

* Difference indicates greater reaction in Spanish-Anglo situations.

It is noted, with reference to the foregoing table:

1) The four-year Anglo is more biased in his attitude toward the Spanish than is the Spanish of the same age group toward the Anglo. It is apparent that the Anglo's concept of the Spanish becomes negative earlier in life than the Spanish's toward the Anglo. This is concluded though it is realized that the average Anglo child of four years has had less intercourse with Spanish than the average Spanish of that age has had with Anglos. Evidently, the exaggerated attitude on the part of the Anglo with little experience concerning the Spanish has arisen from some source other than direct contact.

2) Though the four-year Anglo was more prejudiced in his attitude toward the Spanish than the Spanish at that age level was toward the Anglo, little development between the fourth year and the eighth year was observed in Anglo attitude. This finding was in counterdistinction to the Spanish developmental curves which noted their greatest increment during this period.

3) The Anglo critical ratios between same and not same situations was greater at the twelve-year level than the Spanish (though not significantly so—C.R. = 1.211, significant at .20 level).

Further analysis was possible only after all categories had been analyzed.

Table V represents differences in reaction to Anglo-Anglo and Spanish-Anglo stimuli with respect to neutral and equal categories. As previously pointed out, equality and neutrality are manifestations of lack of emotional arousal and bias.

TABLE V.—SIGNIFICANCE OF DIFFERENCE BETWEEN ANGLO-ANGLO AND SPANISH-ANGLO RESPONSES USING THIRTY ANGLO SUBJECTS IN EACH AGE GROUP ON THE PROJECTIVE TEST OF RACIAL ATTITUDES*

Age Levels	Eff. of Env. Neutral	Adeq. of Pr. Char. Equal	Ending Hero Indeterminate	Ending Theme Indeterminate
12	3.32	10.87	9.53	1.54
8	.79	5.60	5.09	0
4	1.55	2.68	2.88	.38

* Difference indicates greater reaction in Anglo-Anglo situations.

It may be remarked from the foregoing data that neutral responses elicited from the Anglo group to the Anglo-Anglo situations became progressively weighted in comparison with the Spanish-Anglo situation in the older age levels. This is indicative, as previously suggested, of the increased emotional arousal of the older subjects to the Spanish-Anglo situations depicted.

It is noted that the adequacy of the hero figure, closely paralleled by the ultimate outcome of the hero, is most precipitous in loss of neutral reaction.

Ending Theme, (Indeterminate) is incomplete at the eight-year level because no response was observed in the Anglo-Anglo category though there were seven at the Spanish-Anglo level. In this situation the satisfactory ending (Spanish-Anglo) may be less weighted with emotional factors than the indeterminate category.

Table VI presents the significance of the difference manifested by the Anglo subjects between response to the Anglo-Anglo and response to the Spanish-Anglo stimuli. The categories under consideration are: Effect of Environment, (Helpful); Reaction to Environment, (Satisfactory); Adequacy of Principle Character, (Subordinate); Ending Hero, (Victory); and Ending Theme, (Satisfactory). The difference indicates greater reaction in Anglo-Anglo categories 'Effect of Environment,' 'Reaction to Environment,' and 'Ending Theme.' It indicates greater reaction in Spanish-Anglo categories in 'Adequacy of Principle Character' and 'Ending Hero.'

TABLE VI.—SIGNIFICANCE OF DIFFERENCE BETWEEN ANGLO-ANGLO AND SPANISH-ANGLO RESPONSES USING THIRTY ANGLO SUBJECTS IN EACH AGE GROUP ON THE PROJECTIVE TEST OF RACIAL ATTITUDES

Age Levels	Eff. of Env. Helpful*	Reac. to Env. Satisfactory*	Adeq. of Pr. Char. Subordinate†	Ending Hero Victory†	Ending Theme Satisfactory*
12	3.33	6.16	2.87	7.99	6.64
8	.019	3.04	2.91	3.77	6.42
4	.005	2.07	.01	.02	4.99

* The difference indicates a greater reaction in Anglo-Anglo situations.
† The difference indicates a greater reaction in Spanish-Anglo situations.

It is observed from the data in Table VI that victory and subordination are more prevalent in the Anglo's reaction to Spanish-Anglo situations than to Anglo situations alone. This finding parallels that derived from the Spanish data and the conclusions from these data, that ascendance and submission are both manifestations of prejudice, apply to both ethnic groups.

OLDER COMPARED WITH YOUNGER IN EACH AGE GROUP

It is evident from the foregoing analysis that the formation of attitude does not progress at a uniform rate. The Spanish between four and eight years of age and the Anglos between eight and twelve manifest a marked increase in development over that found in other age levels. Evidently, too, there is a period of acceleration particularly at the Anglo level before the four-year level.

The irregularities of growth manifested in the developmental curves demonstrate the impossibility of utilizing extrapolation for prediction, though interpolation is necessarily and legitimately used. The only technique which could be employed to ascertain attitudinal development below the four-year level was to compare the younger members with the older members in each age group. This was significant only at the four-year level, but the other levels were analyzed, too, to note fluctuation which would aid in interpreting the findings at the early age levels.

The technique employed was to dichotomize each age level of each ethnic group by using as samples the twelve older and the twelve younger subjects. It was realized that the small sample would preclude accentuated fluctuations.

Because of the small sample, inconsistencies were manifest, but it was noted, however inconclusively, that the Spanish and Anglos at the three-and-one-half age level rather consistently elicited less difference between reaction to same group and not *same group stimuli*. This would indicate that, *in one year*, a measurable increment in attitudinal development is manifest in both ethnic groups. The data suggest that the origin of Spanish prejudice appears to be very close to the three-and-one-half age level while the genesis of the Anglo appears somewhat before this time.

SUBJECTS WITH SPANISH PARENT AND ANGLO PARENT

An interesting indication of accentuated prejudice was manifested upon analysis of six subjects who had one Spanish and one Anglo parent. Though the number was small, it was found that the subject's frustration and concomitant aggression in the situations depicted were higher than the total for the specific age group. The general prejudice score, too, was higher toward one ethnic group or the other. An attempt was made to ascertain the factors contributory to swing direction.

At first it was hypothesized that the subject adopted the attitude of the dominant parent. This was borne out by the subjects usually being prejudiced for the ethnic group to which the father belonged. In one situation, however, it was found that two brothers (eight and twelve) who lived together had opposite swings of prejudice. It was tentatively concluded with reference to this one situation, at least, that the subjects with one Anglo and one Spanish parent were prejudiced in the same direction as the parent with whom they identified themselves. This was borne out by subjective evaluation and may possibly explain the rather high relationship manifest between parent dominance and attitude. It was also noted that there was a tendency for the subject who had identified himself with the mother to have the girl as heroine in the boy-girl situations of the Projective Test of Racial Attitudes.

The accentuated aggression pattern might be attributed to tension resulting from conflicting tendencies.

SUMMARY OF FINDINGS IN DEVELOPMENTAL ANALYSIS

Both ethnic groups manifest a clear picture of development of racial attitude from the early age levels to the older. Though the attitudinal development of the two groups varied qualitatively and quantitatively, a parallel growth was still in evidence.

An evaluation of the areas of personality studied demonstrated that each revealed either positive or negative prejudice. Interestingly enough it was ascertained that both superiority and submission, defeat and victory are expressions of prejudice—different patterns of attempt to adjust to conflict.

Qualitatively analysis at the twelve-year levels does not reveal a marked difference. Quantitatively, however, the Anglos appear more prejudiced than the Spanish though not significantly so (critical ratio at approximately .25 level of significance).

It is interesting to compare quantitatively the differences in attitudinal growth between the Spanish and Anglos at the selected age levels. The Spanish at the four-year level appeared less prejudiced than the Anglos of the same age (significance of the difference at .25 level), but during the ensuing four years this negative attitude developed to a level approximating that of the twelve-year-olds of that ethnic group. In distinction to this sequence of growth, the Anglo attitude developed little between the fourth and eighth years. Rapid acceleration, however, during the next four years (corresponding in developmental progress to the Spanish four- to eight-year period) put the Anglo at the highest level of prejudice of all groups and ages studied.

Qualitative analysis revealed features of attitudinal development for the two ethnic groups which demonstrated something of the dynamics of personality adjustment and the rôle of prejudice in its function. It was previously mentioned in this study that the Anglos at the four-year level had less acquaintance with Spanish persons than the Spanish. The hypothesis was forwarded for explanation of this factor that some vicarious experience with suggestive qualities had colored the Anglo attitude toward the Spanish.

Analysis of each category and the significance of age on its development brings out some important differences between the Spanish and Anglo groups. The Anglo group showed the greatest development in the final dominance of the Anglo over the Spanish and in the superiority of the Anglo to the Spanish. Contrary to this, the Spanish displayed as the most weighted factor in the total prejudice score the manifestations of frustration in reacting to the Anglo environment and in ultimate victory of the Spanish. Evidently, the Spanish were less assured of their superiority than were the Anglo and used aggression as an adjustment attempt. This finding was supported by subjective evaluation.

In both the Spanish and Anglo groups, defeat of the character with whom the subject identified himself in the stories was a manifestation of prejudice since it occurred more frequently in the Spanish-Anglo situations than in the same group situations. The origin of its development, however, is little understood. In both ethnic groups it appeared as an almost constant positive factor throughout the age levels. Evidently, the genesis of this factor was at a very early age and it reached the twelve-year maximum before the age of four.

The data indicated that the Anglos were less optimistic about the Spanish-Anglo relationship than were the Spanish. The Anglos appeared to be the aggressors and the Spanish merely attempting to adjust to this aggression. The comparison of satisfactory and unsatisfactory endings of the themes, the early development of Anglo attitude, superiority and different rates of prejudice development would demonstrate this. These criteria indicated how the instilled attitude of one group may contribute toward the attitudinal development of the other.

The younger subjects at the given age levels were analyzed independently and compared with the older subjects from the identical age group to ascertain the effect of one year on attitudinal development. The results indicated that, primarily at the four-year level, a measurable increment in prejudice was manifested in both ethnic groups. The data suggested that the origin of Spanish prejudice appeared to be very close to the three-and-one-half-year level while the genesis of the Anglo appeared somewhat before this time.

An interesting indication of accentuated prejudice was manifested upon analysis of six subjects who had one Spanish and one Anglo parent. Though the number was small, it was found that the subject's frustration and concomitant aggression in the situations depicted were higher than the total for the specific age group. The general prejudice score, too, was higher toward one ethnic group or the other. An attempt was made to ascertain the factors contributory to swing direction.

In one situation, it was found that two brothers (eight and twelve years of age) who lived together had opposite swings of prejudice. It was tentatively concluded with reference to this one situation, at least, that the subjects with one Anglo and one Spanish parent were prejudiced in the same direction as the parent with whom they identified themselves. This was borne out by subjective evaluation and may possibly explain the rather high relationship manifested between parent dominance and attitude. It was also noted that there was a tendency for the subject who had identified himself with the mother to have the girl as heroine in the boy-girl situations of the Projective Test of Racial Attitudes.

The accentuated aggression pattern might be attributed to tension resulting from conflicting tendencies.

REFERENCE

1) Franz, Boaz. *Race and Democratic Society.* New York: J. J. Augustin, 1946, p. 27.

VERBAL INTELLIGENCE AND EFFECTIVENESS OF PARTICIPATION IN GROUP DISCUSSION*

NORMAN E. GREEN
Major, USAF, Craig Air Force Base, Selma, Alabama

This report is a part of a broad study of the factor of verbal intelligence or vocabulary power in Air Force teacher training. The investigation is being made at the Academic Instructor Division in the USAF Special Staff School of the Air University where a six-week instructor training course is presented. The test groups for the over-all study were two classes of Air ROTC instructor trainees, totaling two hundred fifty-two members. The population for this part of the study was one class of one hundred twelve of the Air ROTC instructor trainees.

Because conference methods of instruction were being taught, it was felt that it would be valuable to identify some of the factors which contribute to effectiveness of participation. Since it seems logical that vocabulary power might have some bearing on success of conferees, this factor was investigated. In addition, as mentioned below, instruction was carried out in small groups throughout the course, and the opinion was that for purposes of this school, the groups should be arranged heterogeneously according to indicated effectiveness or capacity to succeed. Furthermore, an organismic interpretation of qualitative learning in coöperative group activities demands a heterogeneous arrangement of experiences. Such a situation gives potential basis for optimum interaction and individual growth. Thus, if verbal intelligence could be found to have relationship to effectiveness of participation, it would have value as a criterion for arranging the groups in a manner which tends toward maximum learning. The problem represents an attempt to isolate and describe the relationship of one factor within a 'field' surrounding an organic group learning process.

The major part of the instruction and all of the practice-teaching exercises within the Academic Instructor Course are conducted in groups of about ten student-instructors. In the unit of instruction called Conference Methods of Teaching, with

---

* The author wishes to thank Dr. R. L. Thorndike, Dr. Joseph E. Barmack and Dr. Herman Feifel for their advice and constructive criticisms.

which this paper is concerned, there were seven groups of ten members each and two groups of nine members each. These groups were in session for approximately ten hours of problem-solving discussion centering about such issues (chosen by the students) as, "What Can Be Done to Improve the Airman Information Program?"; "How can We Curb Sex Laxity in American Colleges?"; "What Improvements Can We Make in Evaluation of Officer Effectiveness?" During the ten hours, each group was under the guidance and observation of one faculty-adviser for the first half of the session and another faculty-adviser for the last half of the session.

The procedure, then, was to organize each one of the nine groups of conferees for the ten hours of problem-solving discussion heterogeneously according to scores obtained on two verbal intelligence tests. The tests used were twenty-word, steeply-graded, multiple-choice type. The words were taken from the I. E. R. Intelligence Scale CAVD. These were the same instruments reported on by Thorndike and Gallup in a survey of the American voting population,[1] and earlier by Thorndike alone.[2] This system of determining group membership, having a balanced number of high, intermediate, and low scores in each group, thus equalized the groups as far as this yardstick of verbal intelligence was concerned.

There were twenty-four judges involved in the experiment. These were regular staff members of the school. Each one had had considerable experience with conference methods of instruction, both as participant and as critic teacher. Two observed each group. During the time they were observing the groups in discussion, the faculty-advisers did not know that they would later be asked to give their opinions about the effectiveness of the conferees. Nor did they have knowledge of the vocabulary scores received by the group members.

On the day when this discussion phase was completed, each instructor involved was given a written memorandum on which he was asked to submit the names of three student-instructors whom he considered to be most effective as conferees during the periods he observed the group. In order to determine this judgment, they were asked to concentrate upon those who participated most frequently and at the same time gave good quality contributions. They were to name these three conferees in order of effectiveness. In addition, they were asked to submit the names of three conferees whom they considered to be least effective. Reticence was the principal criterion for determining this. The three least effective were also to be named in numerical order. It is the writer's opinion that, in view of the extended experience of the faculty-advisers with this conference method, the requested criterion of effectiveness of participation was a sound one. The judges were particularly aware of the dangers which develop through dominance of the discussion by one or a few members. And they recognized the benefits of constructive contributions, frequently termed 'committee work.'

---

[1] "Verbal Intelligence of the American Adult," R. L. Thorndike and G. H. Gallup, Journal of General Psychology, Vol. 30, Jan., 1944.

[2] "Two Screening Tests of Intelligence," R. L. Thorndike, Journal of Applied Psychology, Volume 26, 1942.

When the names of the most effective and least effective participants in each of the twelve groups were received, a point scoring system was established. A conferee named most effective of the entire group received a score of plus 3; one named second most effective received a score of plus 2; and the one named third most effective of each group received a score of plus 1. The same scoring was used for the least effective members, except that the figures were negative. Thus, if a conferee was named by each of the two judges (who worked independently) as the most effective of a particular group, he would receive the highest possible score, a plus 6.

This scoring system, later used in the refinement of the data, permitted an equitable discrimination among effectiveness in each group. It was thus possible to (1) pick out the three most effective in each discussion group, (2) pick out the three least effective in each group, (3) pick out the top most effective participant in each group, (4) pick out the one representing least effectiveness in each group, (5) arrange the members of each group in rank order of effectiveness, and (6) through an adjustment of the raw scores, to compute a product-moment correlation coefficient between effectiveness and verbal intelligence for the whole group of fifty-four participants who were named on the effectiveness scale.

From this point on, the experiment and treatment of the data were further organized into two phases. The first was an attempt to establish a statistical relationship between verbal intelligence and effectiveness of conferees. This included the assemblying of the vocabulary scores of those named in the top bracket of effectiveness in all the groups (N = 27) and all those named in the least effective bracket (N = 27). It also included assemblying the vocabulary scores of the nine conferees named first in effectiveness of participation and the scores of those nine conferees who were named as representing the least effective member of each group.

Table I presents the statistics for two larger groups, and shows the significance of the difference between the two means.

TABLE I.—SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE MEAN VOCABULARY SCORE OF THE MOST EFFECTIVE PARTICIPANTS IN GROUP DISCUSSION AND THE MEAN VOCABULARY SCORE OF THE LEAST EFFECTIVE PARTICIPANTS

	N	Mean Vocab. Score	Standard Deviation
Most Effective (X)	27	29.77	4.4
Least Effective (Y)	27	26.29	3.9
$\bar{X} - \bar{Y}$		3.48	
Standard Error of Difference		1.14	
Critical Ratio		3.05	

The null hypotheses, that there is no true difference in verbal intelligence as measured by this yardstick between the most effective and least effective participants, is rejected at the .01 level of significance.

In dealing with the smaller groups in a similar manner, the small sample theory and formula were employed in the refinement of the data to determine the statistical significance of the actual difference of six points between the means of the two groups as shown in Table II.

In this case, the difference is significant at the .03 level. Results obtained from the two treatments of the data show that a true difference exists in the groups and the difference is not due to chance.

Thus, from the first phase of the study it was evident that a

TABLE II.—SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE MEAN VOCABULARY SCORE OF THE TOPMOST EFFECTIVE PARTICIPANTS OF EACH DISCUSSION GROUP AND THE MEAN VOCABULARY SCORE OF THE PARTICIPANTS SHOWING MINIMUM EFFECTIVENESS

	N	Mean Vocab. Score	Standard Deviation
Participants with Maximum Effectiveness (X)	9	31.66	4.4
Participants with Minimum Effectiveness (Y)	9	25.66	5.3
$\bar{X} - \bar{Y}$		6.0	
Standard Error of Difference		2.43	
Critical Ratio		2.88	

statistical relationship existed. In phase two, an attempt was made to discover more about the degree of relationship.

Table III represents a breakdown of each of the nine groups showing that within each (except group #9), there was a difference in 'vocabulary power' consistently in favor of those named as effective participants. The range of the differences is from zero difference in group 9 to a difference of 7.67 points in group 2. The mean difference is 3.48.

TABLE III.—DIFFERENCE IN MEAN VOCABULARY SCORES OF MOST EFFECTIVE AND LEAST EFFECTIVE PARTICIPANTS IN NINE DISCUSSION GROUPS

Group	1	2	3	4	5	6	7	8	9
Mean Vocab. Score of most effective	31.67	32.33	31.67	30.66	29.00	28.33	27.66	30.33	26.33
Mean Vocab. Score of Least Effective	27.33	24.66	28.00	24.66	24.00	27.66	27.00	27.00	26.33
Difference	4.34	7.67	3.67	6.00	5.00	0.67	0.66	3.33	0.0

Table IV presents, for each discussion group, the rank-difference correlation coefficient between the members' effectiveness scores and their scores in verbal intelligence. All groups had ten members except groups 7 and 9, which had nine members each. Although all coefficients are positive in a range from

TABLE IV.—RANK DIFFERENCE CORRELATION COEFFICIENTS BETWEEN VERBAL INTELLIGENCE AND EFFECTIVENESS OF PARTICIPATION FOR NINE DISCUSSION GROUPS

Group	1	2	3	4	5	6	7	8	9
Rank Diff. Correlated Coefficient	+.29	+.57	+.36	+.40	+.52	+.10	+.27	+.30	+.09

+.09 to +.57 (mean coefficient is +.32) the small samples involved precluded any proof of statistical significance.

In the final treatment of the data, the effectiveness scores of all fifty-four participants representing the top and bottom brackets of each group were paired with the vocabulary scores in a computation of product-moment correlation. This coefficient was found to be +.45, with a standard error of .10, representing a correlation value significantly larger than zero at the .01 level.

Although these correlations are not of high order, they indicate that a positive relationship is present to at least a fair degree. A comprehensive interpretation would not fail to take cognizance of other seemingly obvious factors having influence on one's ability to participate effectively. Some of these might be familiarity with the topic of discussion, interpersonal relationships in the group, motivation and general interest. If these factors also are of importance, then the consistent positive statistics in favor of verbal intelligence may give greater reason for attaching a relative value to this latter variable as one concomitant.

These results might then be construed to demonstrate that verbal intelligence was a contributing factor to effective participation. Practically speaking, this means that the administration of these instruments and the use of the scores may serve as a valuable aid in setting up small learning groups balanced on the basis of the ability of the conferees to participate. A continuation of the investigation of effectiveness in group processes must be directed toward study of other probable attending factors indicated above. Beyond this, further research may reveal additional educationally significant relationships between scores on these tests of verbal intelligence and capacity to perform successfully in other teaching-learning practices.

BOOK REVIEWS

WILLARD L. VALENTINE AND DELOS D. WICKENS. *Experimental Foundations of General Psychology.* Third Ed. New York: Rinehart and Co., 1940, pp. 472.

This third edition is a rather thorough revision of a text which has had wide usage since the appearance of the first edition. Nevertheless, the general pattern of the text is the same. In each chapter there is a preliminary orientation to the subject matter and problems plus definitions and in some cases methodology. This is followed by selected investigations in condensed form which usually include introduction and hypothesis, method, results and a summary of interpretations. There is also a summary of the chapter. Throughout the book there is some emphasis upon correcting misconceptions about the nature of psychology, and the inclusion of materials which have a part in "determining the theoretical structure of modern psychology."

In this revision there is a new chapter on personality. Most of the material on perception is new, and the chapters on conditioning and learning are drastically revised. In most of the other chapters new experiments have been added or substituted for material omitted. In all the revision the selection of material and discussions have been centered on the more recent trends in the field. In particular there are new materials on social reactions and reports of investigations done by military psychologists.

It is stated that the book is "concerned principally with psychological facts and principles on which there is a majority, if not a universal, agreement." Some sections found in psychology texts are not included here. No particular system of psychology is emphasized. There is, however, emphasis upon utility and applications. In general, the materials indicate how a psychologist collects information he needs and how he interprets the data collected. The book is designed as a supplement to standard texts.

Most readers will agree that the selection of experiments has been done with discrimination. Nevertheless some may consider that in too many instances animal rather than human experiments were chosen. Furthermore, if contemporary trends are to be emphasized, the material on perception could have been both broader and more extensive.

The material throughout the book is interesting and is concisely and clearly presented. It would appear that the avowed aims have been adequately achieved. It is noteworthy that the figures and typography are excellent. There can be little doubt that this revision will be popular and enjoy wide usage.

MILES A. TINKER

University of Minnesota

SUSAN DERI. *Introduction to the Szondi Test.* New York: Grune and Stratton, 1949, pp. 354.

Mrs. Deri's familiarity with the Szondi test extends over a period of about eleven years, the first four of which were spent working closely with Dr. Lipot Szondi in the early applications of this method to personality diagnosis. Dr. Szondi's highly laudatory Foreword is confirmation of the authority with which the author writes.

The Szondi test consists of forty-eight photographs of mental patients. These fall into six sets each containing eight pictures; one each of a homosexual, a sadist, an epileptic, an hysteric, a catatonic schizophrenic, a paranoid schizophrenic, a manic depressive depressive, and a manic depressive manic. One set of photographs is displayed to the subject at a time. From each set the subject selects the two pictures he likes most and the two he dislikes most. When this process has been completed, the subject chooses from among the twelve 'most liked' the four for which he has the greatest liking and from among the twelve disliked the four arousing the greatest aversion. These responses are transferred to an appropriate graphic record sheet. The test must be repeated "at least six, preferably ten, times, with at least one day intervals between administrations, to be able to give a valid clinical interpretation of the personality." An optional extension, entitled the 'factorial association experiment,' involves asking the subject to tell stories about each of the eight pictures finally chosen.

The major portion of the book is devoted to an exposition of the interpretation of scores. Need-tension is regarded as the driving force "in the sense of directing the person to perform certain acts and to choose or avoid certain objects." The type of activity or object depends upon the kind of need. "Its field of application is again similar to that of other projective technics; in other words, as a diagnostic instrument for clinical use or for the interpretation of the so-called normal personality, vocational guidance, experimental social psychology and a variety of fields of research."

Mrs. Deri avoids an expression of attitude toward the gene-theory which this test was originally intended to 'prove experimentally.' *According to Szondi's theory*, "the mental disorders represented in the test are of genetic origin and the subject's emotional reactions to these photographs were believed to depend upon some sort of similarity between the gene-structure of the patient represented by the photograph and that of the subject reacting to the photograph." Mrs. Deri takes the position that the test's effectiveness has been established whether or not the early hypothesis is accepted.

To the reader who is accustomed to look for experimental data in support of claims of usefulness or validity, this volume is wholly inadequate. Elaborate explanations of the significance of various factors are given without even frequency tables. Sums, differences and ratios are computed and interpreted without thought for the problems of reliability involved. Shifts in type preference or aversion from one day to the next are regarded as indications of changes in the needs and tensions of the subject rather than errors of measurement. Validity in the usual sense appears to be dismissed in the following paragraph.

"The superficial appearance of normalcy is responsible for the extreme difficulties inherent in the problem of validating studies on the basis of observable behavior or verbal or written questionnaires. Many basically unhappy individuals who are unable really to become emotionally attached to any person or object would rate extremely high on a written adjustment inventory, or on the basis of observation."

This reviewer can only conclude that, on the basis of Mrs. Deri's book, the Szondi test is one more unproven instrument proposed as an aid in personality diagnosis. Pending the publication of adequate supporting evidence, the psychologist using this method is proceeding on the basis of faith and should be willing to recognize the peril of his course.

George K. Bennett

The Psychological Corporation

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 41 December, 1950 Number 8

THE DEVELOPMENT OF A FUNCTIONAL COURSE IN EDUCATIONAL PSYCHOLOGY FOR TEACHERS[1]

HERBERT KLAUSMEIER
Colorado State College of Education
and
DONOVAN SWANSON
Chico State College

As Trow has pointed out in an earlier issue of this JOURNAL: "A visit to a school is still a saddening experience to an educational psychologist. With due respect to the many exceptions, so many things are wrong, so completely wrong, that the visitor wonders what has become of all the well-established 'principles' that the teachers and administrators supposedly learned in order to pass the psychology courses required for their certification."[2] He suggests that improvements can and should be made at the source; namely, in organizing and teaching the courses. The purpose of this paper is to present an account of a study to develop a functional course in educational psychology for teachers.

PRELIMINARY CONSIDERATIONS

Under the leadership of Professor Lucien B. Kinney, a comprehensive curriculum development and evaluation program in teacher education was begun in 1947 at Stanford University.

---

[1] Herbert J. Klausmeier and Donovan A. Swanson. *The Development of a Functional Course in Educational Psychology for Teachers.* Unpublished Ed.D. Dissertation, Stanford University, 1949. The writers are indebted to Professor Lucien B. Kinney, who served as chairman of the thesis committee.

[2] William C. Trow. "How Educational Psychology and Child Development Can Contribute to the Preparation of Teachers." *The Journal of Educational Psychology.* 39:129, March, 1948.

As part of this program, the role of educational psychology in the professional sequence was studied. The foundations for the study were laid in the spring quarter of 1948 when Professor Kinney taught the course and the authors were assistants in education. The evaluation was carried on during the autumn and winter quarters of 1948.

This study was directed to the organization and evaluation of a course in educational psychology designed for one quarter's work, carrying four quarter-hours of credit, required of all students seeking elementary- or secondary-school teaching credentials at Stanford University. The class population for the autumn quarter consisted of one hundred sixteen students; for the winter quarter, seventy-seven students. For both quarters approximately one-half had graduate standing; the remainder were undergraduates. All students had completed a course in general psychology; in addition, elementary-school candidates had completed a course in child psychology. In setting up the study the following major problems were defined and analyzed:

1) What are the competences of effective classroom teachers? An explicit definition of what constitutes teaching competence is essential in building a professional curriculum in teacher education. Unless the behaviors of effective teachers are operationally defined in terms of teaching competence, courses cannot be organized with any degree of validity, nor can the effectiveness of such courses be determined. The "Factors in Teaching Competence" as defined by the California Council on Teacher Education is a comprehensive statement of teaching competence upon which the professional program in teacher education at Stanford University is built.[3]

2) Which teacher competences should be developed in educational psychology as part of the professional sequence? The "Factors in Teaching Competence" provides the objectives to be developed in the total professional sequence. The responsibility of each of the various courses was assigned in staff conferences as the sequence of courses was established. In this manner, integration is provided for and duplications and omissions are avoided.

---

[3] Committee on Definition of Teaching Competence. California Council on Teacher Education. "Factors in Teaching Competence." Mimeograph statement. Stanford University, July, 1948.

3) In the pre-service education of teachers, what experiences should be provided in educational psychology to develop the competences? Learning experiences provided should: (a) utilize the unique contributions of the field of educational psychology to effective teaching and incorporate the objectives of the course, (b) be organized according to a developmental sequence which illustrates psychological principles of effective teaching, (c) be selected to illustrate the activities of the classroom teachers, and (d) utilize the resources of the public schools and the community.

4) How effective are the experiences in developing teacher competence? Evaluation is essential in curriculum development in order to determine (a) which competences are developed and to what extent, (b) which experiences are valuable in developing competences and to what extent, (c) which procedures are effective, and (d) what are the implications for practice and improvement.

Basic to this study was the assumption that teacher education is primarily an area of personnel development of which scholarship is one aspect. Thus, educational psychology is defined as an area of personnel development within the field of teacher education in which resources and procedures from general psychology, human biology, and cultural anthropology are brought to bear upon classroom problems. A second assumption was that "Factors in Teaching Competence" as defined by the California Council on Teacher Education represents a workable hypothesis on which to build a professional program in teacher education. Therefore, the objectives for educational psychology can be drawn from the "Factors in Teaching Competence" and defined operationally in terms of teaching competence.

To summarize, in setting up the program in educational psychology, this sequence was followed: (1) objectives in the course were established in terms of student behavior indicative of teaching competence; (2) learning experiences were tentatively organized to achieve the objectives; (3) sources of evidence to be used in observing the behaviors were identified; (4) evaluation instruments and procedures were developed for obtaining data; (5) the data were gathered, summarized, and interpreted to make the best possible judgment concerning student attainment of the objectives during the autumn quarter; (6) on the basis of the evaluation during the autumn quarter, certain revisions were made in objectives, course experiences, and evaluation instruments; (7) data were gathered, summarized, and interpreted during the winter quarter; (8) on the basis of the evaluation, autumn and winter, conclusions were drawn and recommendations made.

DEFINING OBJECTIVES IN EDUCATIONAL PSYCHOLOGY

Agreement was reached that the primary responsibility of educational psychology is to develop the competence necessary for the prospective teacher to organize learning activities in the classroom in which the teacher:

1) Adapts psychological principles of learning to individuals and groups. (a) Provides for individual capacities of pupils. (b) Employs adequate motivational procedures. (c) Organizes learning experiences which possess meaning and structure for the pupil. (d) Uses effective practice procedures. (e) Selects and appraises experiences in terms of their transfer value.

2) Adapts psychological principles of growth and development to individuals and groups. (a) Interprets pupil behavior in relation to his present situation as shaped by past experience and as influenced by future expectations. (b) Adapts classroom experiences to the social backgrounds of the pupils. (c) Integrates classroom experiences with life in the home. (d) Integrates the social dynamics of the classroom. (e) Develops in the classroom an atmosphere conducive to mental health.

3) Adapts psychological principles of counseling and guidance to individuals and groups. (a) Knows each pupil as an individual. (b) Recognizes range of emotional maladjustments. (c) Provides success experiences for all pupils.

4) Adapts psychological principles of evaluation and measurement to individuals and groups. (a) Uses appropriate techniques for evaluating the effectiveness of school learning. (b) Uses appropriate techniques for measuring achievement. (c) Uses grading system with clearly defined purposes.

"Develops professional behavior" is a fifth major objective shared with all other professional courses.

After these general objectives had been agreed upon, specific aims for the course were developed. It was necessary that these aims be sufficiently complete and definite to orient the course experiences, to direct the selection of resources and materials, and to serve as the basis for evaluation. The criteria for specific aims were adapted from those developed by Kinney.[4]

A. *Validity.*—Can the aims be justified on the bases of (1) social and professional needs of the students, (2) personal needs of the students, and (3) feasibility as determined by the possibility of attaining them?

The pioneer work of the California Council may be considered as having tentatively established the social and professional validity of these aims. Further validation has been obtained from the considerable body of research already carried out at Stanford University. While the personal needs of the students may be assumed to follow as a matter of course from the social and professional needs, the psychology of learning in general and of motivation in particular requires that the students accept the personal significance of the aims.

Another important question in regard to validation of an aim is whether it can be achieved with reasonable expenditure of time and effort. The answer here can only be tentative. Only through experience in teaching the course can the feasibility of an aim be finally established.

B. *Comprehensiveness and Selectivity.*—Do the aims include the outcomes of major importance and represent the unique contributions of the field, and differentiate the more important from the less important?

Comprehensiveness and selectivity of the aims was established on the basis of expert opinion and related research.

C. *Suitability of Form.*—The final consideration was to state the aims in a form which would be useful as a basis for teaching and evaluation. Expressing aims in terms of desired behavior has many advantages—it is simple, and it avoids verbalism, as well as the pitfalls of expressing outcomes in terms of traits. The criteria for form are:

1) Can I recognize this competence in a student?

2) Can I differentiate a student who has the competence from one who does not?

3) What kinds of behavior will distinguish the student who has the competence from one who does not have it?

---

[4] Lucien B. Kinney. "Criteria for Aims in Mathematics." *The Mathematics Teacher.* 41:99–103, March, 1948.

To take an illustration: Most educators agree that an important aim in educational psychology is to develop prospective teachers who are competent to appraise the readiness of a pupil to learn a specific ability or concept. Stated operationally, the prospective teacher who is competent in this area demonstrates the following behaviors:

1) Understands the major factors which determine readiness to learn.

2) Recognizes how present interest may be used to motivate classroom learning.

3) Describes suitable diagnostic techniques for determining readiness to learn a particular ability or concept.

4) Evaluates diagnostic procedures used in the classroom.

5) Plans learning experiences which take into account factors of readiness.

This statement of operational definitions illustrates the extent to which one sub-section of aims under "Provides for individual capacities of pupils" was delimited and defined. It also indicates the kinds of experiences necessary to develop such competence and suggests areas in which student behavior may be observed and evaluated. Readings, group discussions, classroom visitations, films and resource personnel, and lectures all contribute to the development of these behaviors. Making an observational case study of a school child also contributes to competence in this area. The actual planning of learning experiences which take into account factors of readiness is suggested by the nature of the operational definitions.

DEVELOPING THE COURSE ORGANIZATION

The over-all pattern of the course comprised five major areas:

1) What contributions has educational psychology to successful teaching?

2) What is the typical sequence and pattern of meaningful learning?

3) What factors determine the social climate of the classroom? What is the significance of these factors for the teacher as an organizer of learning?

4) How may the teacher use the resources of the school to facilitate the adjustment of the individual child?

5) How may the teacher evaluate the effectiveness of his work?

The procedures employed in developing teaching competence as related to the first three major areas are illustrated below:

1. What contributions has educational psychology made to successful teaching? (a) What is the nature of the field? (b) What contributions do the authors of the textbooks claim for it? (c) With what kinds of classroom problems is educational psychology concerned? (d) What preparation do I as a teacher need in this field?

The purposes of this unit were: (a) to orient the student to the rôle of the teacher as an organizer of learning experience, (b) to provide a variety of motivational experiences for the major problems to be considered in this course, (c) to ensure student participation and acceptance of the objectives, and (d) to determine student interest and needs.

The students sought answers to the above problems in various activities. They scanned the textbooks. They visited classrooms in the public schools where they noted problems of organizing learning and problems of classroom control which require the use of principles of educational psychology. Information from these sources was brought together in small discussion groups which met each week during the quarter. In the first group meetings the students, on the basis of their experiences and with the guidance of the instructor, considered: (a) What topics should be dealt with in the course? (b) What outcomes sought? (c) How should the outcomes be evaluated?

From observations of public school classrooms the students derived a delineation of the scope and contributions of educational psychology to effective teaching practices. They observed that the organization of learning experiences is the central activity of the teacher; and it was in the rôle of organizing learning experiences that they, as prospective members of the profession, pictured themselves. Questions suggested for exploration and major consideration were included in the next unit.

2. What is the typical sequence and pattern of meaningful learning? (a) How may the teacher provide for individual capacities of pupils? (b) How may the teacher determine readiness to learn a particular ability or concept? (c) How are

children motivated to learn? 1) How may the teacher assist pupils to establish realistic goals? 2) What are the effects of rewards and punishments? 3) How is knowledge of progress assured? (d) How may the teacher organize learning experiences which possess meaning and structure for the pupils? (e) What are effective practice procedures? (f) How may the teacher secure transfer?

The common sequence in human learning involves a gradual transition from immature reactions to mature methods of responding and learning. During this period of development, the student, through active experience involving exploration, discovery, and verification, acquires meanings and generalizations which lead to competence. A review of procedures followed in teaching Topic II, "What is the Typical Pattern and Sequence of Meaningful Learning?" illustrates "the developmental sequence of learning."[5]

The major purpose of this topic was to develop the competence necessary for the prospective teacher to organize learning experiences in the classroom according to psychological principles of learning.

Step 1.—To achieve the above purpose, a variety of concrete experiences was provided. The teaching process was presented in a variety of settings with emphasis on the situation. A classroom visit was made in which the student observed a learning situation. A film was presented and discussed which illustrated typical classroom activities. Students read general and special methods books. Lectures were used to illustrate various teaching procedures. Class discussions were held centering on teaching practices in specific subject-matter areas. Resource speakers described teaching procedures in special fields. In small group meetings, students discussed procedures with which they were now familiar. Thus, the situations in which the concepts were experienced were simple, numerous, and varied so that the concepts remained the common elements. In the lectures and the group meetings the concepts acquired increasingly abstract meaning as opportunity was provided to differentiate out the common element. It was not considered sufficient to have the concepts explained in words. . When this occurs, too often the result is a verbalism rather than a usable concept.

---

[5] Lucien B. Kinney. "The Operational Plan in the Classroom." *School and Society.* 28:145-149, September, 1948.

Step 2.—The concepts which had been acquired as a result of the numerous, varied experiences and had become increasingly abstract through differentiation were then manipulated on an abstract or symbolic level and relationships were explored. Opportunity to manipulate the acquired concepts was provided in class discussion. The lectures at this point were designed to help the students explore relationships by providing the necessary clues for the development of the group project.

The purpose of the group project was to provide a focus for the integration of concepts and understandings relative to how the pupil learns and how the teacher organizes learning experiences to facilitate the pupil's learning. In developing the learning unit, the students manipulated concepts and related them to a meaningful pattern. It was considered important that the pattern of concepts or generalizations should represent the students' own generalizations, discovered by them. The instructor planned the situation and provided clues; but the exploration, discovery, and verification represented the learning process of the student.

To illustrate the above further, the concept of the 'whole child' and the generalization, 'learning is purposeful,' are not to be memorized but are to be discovered in a variety of situations. The vocabulary of educational psychology must depend for adequacy and richness of meaning on the significance and variety of experiences that have contributed to the meaning. Competences to be developed through educational psychology are not to be memorized as verbalisms, but are to be developed as part of the individual student's behavior in meeting problem situations.

The necessity for concrete and numerous experiences is clear. It should be recognized, however, that the mere accumulation of experiences is of little value until they are organized. At the symbolic level the student is able to manipulate his concepts, classifying them by using one word or symbol to represent many items with a common characteristic, and relating them to other concepts and behaviors so as to form a part of a total pattern. Students, in developing the learning unit, related the procedures of teaching which they had seen, heard of, and read about, with the planning of experiences for a particular classroom situation. The students thus carried over their understanding from old to new processes, and saw the place of the new in the old pattern.

Step 2 has for its purpose, then, the transfer of the students' attention from the situations in which the concept or process was being studied, to the concept or process itself, in order that it may be manipulated and understood in the context of the system of educational practice where it belongs. It then becomes a generalized and symbolized idea with which to interpret and control new situations.

Step 3.—Learning is not meaningful or useful unless the student can utilize in new situations what has been learned previously. The concepts, processes, and principles learned must be available in future teaching situations, the characteristics of which cannot be anticipated with exactness. The final step in the learning process, therefore, must be to identify if possible the utilization of the concept in new situations—to identify the characteristics of those situations in which it is appropriate.

In general we may think of this stage as one during which the newly defined process or concept is integrated with the professional work of the student. Left to itself this integration may or may not take place. 'Transfer of training' may occur by accident, but it is not part of good teaching to leave the final result to chance.

To ensure transfer and to develop planning competence, the group project involved the organizing of learning experiences suitable for a hypothetical, but typical, classroom situation. The development of the learning unit required that the students participate in a discussion group resembling a teachers' curriculum-planning committee. The integration of concepts and generalizations, acquired previously, as described in Steps 1 and 2, was accomplished in the final preparation of the written learning unit. Thus, in Step 3, generalized and organized procedures were carried back to significant problems.

The sequence of the learning process begins and ends in life situations. In between is a stage of abstraction, organization, and generalization. But, throughout, effective learning at the college level, as elsewhere, depends on the availability of many resources beyond the textbook and the lecture room—classrooms for observing teachers in action, demonstrations of effective teaching, group discussions, dramatizations, resource personnel, children available for study, and other means by which the teacher education institution and public school system are brought in close relationship.

The teacher, while directing experiences in the classroom according to psychological principles of learning, becomes concerned with (a) the wide range of developmental patterns of behaviors present and (b) the kind of persons the pupils are becoming. These problems are dealt with in the next unit.

3. What factors determine the social climate of the classroom? What is the significance of these factors for the teacher as an organizer of learning? (1) What should the teacher know about growth, development, and behavior? (2) What should the teacher know about the individual child? (3) What should the teacher know about the child's family? (4) What are the dynamics of the peer culture? (5) What is the rôle of the teacher in the democratically constructive school? (6) What general principles concerning teacher behavior evolve?

The purposes of this unit were: (a) to define the social dynamics that operate both within and without the classroom which influence the course of effective learning, (b) to analyze the factors the teacher must deal with which determine the kind of person the pupil will develop into in the course of his learning, and (c) to achieve the outcomes operationally defined under the major objective—"Adapts psychological principles of growth and development to individuals and groups."

In addition to the experiences provided in the preceding topics, each student made an observational study of a child of school age in the community. The purpose of this study was to provide a focus for the organization and integration of the student's understanding of child behavior and development. The student observed the child's behavior in a variety of situations. He visited the home in an attempt to gain insight into the emotional stability of the home, the pattern of training, and parental expectations. All behavior observed was recorded in the form of anecdotes that described objectively (a) physical development and health status, (b) relationships with parents and siblings, (c) relationships with age mates, (d) school adjustment, and (e) interests and abilities. Then the student formulated a series of hypotheses to account for the particular pattern of behavior and development which had been discovered. These hypotheses were checked against an organized framework of psychological principles. The student described kinds of information, not available, which he recognized as necessary to make a complete diagnosis. Finally, the student suggested a plan of action for helping the child if the data so warranted. The developmental sequence of learning was implemented in this unit with the study of an individual child serving similar purposes as those outlined for developing the learning unit.

EVALUATING OUTCOMES

The general plan of this study was that of professional curriculum development beginning with definition of objectives and concluding with appraisal of outcomes. The evaluation procedures incorporated those which were found effective by other researchers[6] and new methods were devised. As indicated previously, some of the experiences in the course were planned to develop theoretical understanding of a classroom problem; some to develop ability of the student to plan what he would do in this case; and some to develop ability to demonstrate effective behavior in meeting the situation. Therefore, competence of the students was appraised at three levels: theoretical, planning, and performance. Three paper-and-pencil tests were employed to determine competence at the theoretical level; three evaluation forms were constructed to appraise competence at the planning level; two forms were constructed to summarize performance behavior. A self-evaluation form and a student-opinion questionnaire complete the ten major instruments employed in the winter quarter. To determine change in behavior from beginning to end of each quarter, the results of a Principle and Judgment Test, an evaluation of first and last classroom observation report, and an analysis of observer comments and description of student behavior were employed.

Specific techniques utilized to determine the extent to which experiences in this course and not other courses had, for the major part, contributed to the development of the competences were: (1) staff meetings were conducted to define and delimit the objectives to be achieved in this course and other courses in the professional curriculum; (2) conferences were held with individual students from beginning to end of each quarter and data collected; (3) trained observers met with the students once per week and collected data from beginning to end of each quarter; (4) data were collected for two consecutive quarters with two groups of students, (5) tests and evaluation forms were administered at beginning and end of each quarter. A brief description of instruments, representative data on which conclusions were based, and comparisons of data for both quarters follow.

---

[6] M. E. Troyer and C. R. Pace. *Evaluation in Teacher Education.* Washington: American Council on Teacher Education. 1944.

Paper-and-Pencil Test Results.—A Principles and Judgment Test was constructed and administered on the second and the last day of the quarter. The purpose of this test was to appraise student competence at a theoretical level with respect to attainment of two major objectives: "Adapts psychological principles of learning to individuals and groups," and "Adapts psychological principles of measurement and evaluation to individuals and groups." Also, the results were analyzed to ascertain change from beginning to end of quarter.

As indicated in Table 1, the mean score for the first administration was 38.31 with a standard deviation of 5.44 units and a range in score from 25 to 52. The mean score for the second administration was 47.31 with a standard deviation of 5.92 units and a range in score from 27 to 59. On the last administration, three scores were below the mean score of the first administration; in terms of standard deviation, the increase was slightly less than two. The critical ratio of the increase was found to be 9.48, which indicates that the difference was not due to chance.

The Case of Mickey Murphy,[7] a standardized instrument, was administered at the end of the quarter. The purpose of this test was to appraise competence at a theoretical level with respect to the other two major objectives: "Adapts psychological principles of growth and development to individuals and groups" and "Adapts psychological principles of counseling and guidance to individuals and groups." The data are summarized in Table 2.

---

[7] Warren R. Baller. *The Case of Mickey Murphy.* Lincoln: University of Nebraska, 1943.

The mean score for each of six objective-type parts of the test and the mean of the total score was computed for students with teaching experience and students with no teaching experience. The means were converted into percentile norms established by Ballor. The range in percentile rank of mean score for sixty students with no teaching experience was from 62 to 91; the range in raw score was from 145 to 201. The range of percentile rank for twelve students with teaching experience was from 43 to 63; the range of raw score was from 131 to 213. The difference in total mean score between those with experience

TABLE 1.—SCORES ON JUDGMENT AND PRINCIPLES TEST, FIRST AND SECOND ADMINISTRATION, WINTER QUARTER

First Administration		Second Administration	
Score	Frequency	Score	Frequency
		57–59	3
		54–56	6
51–53	1	51–53	11
48–50	1	48–50	23
45–47	5	45–47	10
42–44	17	42–44	7
39–41	12	39–41	9
36–38	16	36–38	1
33–35	10	33–35	1
30–32	7	30–32	0
27–29	2	27–29	1
24–26	2		
Number	73		72
Mean score	38.31		47.31
Standard dev.	5.44		5.92
σ of mean	0.64		0.70
Diff. of means			+9.00
σ of difference			0.95
Critical ratio			9.48
Reliability coefficient			0.74

and those with no experience was 2.83 units. A critical ratio of this difference was found to be 0.35 indicating that the difference was not significant. The mean score of the students with no teaching experience was higher by 4.03 units, winter than autumn. A critical ratio of 1.29 indicates that this difference is not significant, P. being 0.20.

TABLE 2.—MEAN SCORES FOR SIXTY STUDENTS WITH NO TEACHING EXPERIENCE AND TWELVE STUDENTS WITH TEACHING EXPERIENCE ON *Mickey Murphy*, WINTER QUARTER

No Teaching Experience*				Experienced Teachers		
Test Part	Mean Score	Comparable Percentile Rank	Standard Deviation	Mean Score	Comparable Percentile Rank	Standard Deviation
II	19.33	87	3.12	18.67	59	3.62
III	8.22	78	0.98	8.17	55	0.90
V	51.25	91	7.20	50.75	63	11.74
VI	25.46	68	4.06	27.00	55	4.00
IX	47.75	77	6.51	46.00	50	9.00
X	22.55	62	5.38	20.75	43	6.06
Total	174.83		18.63	172.00		25.86

* The percentile norms defined by Baller are for undergraduate students with no teaching experience. In this study, thirty-nine students were undergraduates.

The data obtained through the Principles and Judgment Test and *Mickey Murphy* indicated that the students attained a relatively high degree of competence at the theoretical level with respect to the four major objectives defined for the course. The change from beginning to end of quarter was significant. Marked differences were found among individuals.

Planning Competence.—Behavior at the planning level was observed in written classroom observation reports, in written observational case studies, and in written group projects. Eleven learning units were developed in small group meetings. The purpose of the learning unit was to develop planning competence related to two major objectives: "Adapts psychological principles of learning to individuals and groups" and "Adapts psychological principles of measurement and evaluation to individuals and groups." The criteria employed in rating the units were ten operational definitions of the two objectives. Each of these criteria was rated on a 4, 3, 2, 1 scale. A 4 rating

TABLE 3.—RATING OF ELEVEN WRITTEN GROUP PROJECTS, WINTER QUARTER

Criteria for Rating	Frequency of Rating				Mean Rating
	4	3	2	1	
1) Plans experiences which utilize present interest to focus attention on subject matter.	8	3	0	0	3.73
2) Plans experiences which take into account factors of readiness.	8	2	1	0	3.64
3) Plans experiences which meet the needs of heterogeneous abilities and interests.	6	5	0	0	3.55
4) Plans a variety of experiences to motivate a particular learning.	7	3	1	0	3.55
5) Plans experiences which satisfy pupil needs.	6	4	1	0	3.45
6) Plans experiences which ensure transfer.	7	3	0	1	3.45
7) Plans experiences in which motivation is developed through positive interests.	7	2	1	1	3.36
8) Plans experiences which provide opportunity for successful participation of all pupils.	4	6	1	0	3.27
9) Plans experiences in which appropriate practice procedures are employed.	4	6	1	0	3.27
10) Plans experiences which provide an overview of the whole problem and its related parts.	6	1	4	0	3.18

was interpreted to indicate competence expected of effective classroom teaching, a rating of 3 to indicate competence expected of effective student teaching, a rating of 2 to indicate competence necessary to enter student teaching, and a 1 rating to indicate doubtful competence for student teaching. The results of the rating are summarized in Table 3.

The major purpose in developing the observational case study was to develop planning competence as outlined under the other two major objectives: "Adapts psychological principles of growth and development to individuals and groups" and "Adapts psychological principles of counseling and guidance to individuals and groups." The form developed for evaluating the case study included ten operational definitions as criteria; each rated on a 4, 3, 2, 1 scale as described for the learning unit. The results are summarized in Table 4.

The data thus obtained through evaluation of the learning unit and the observational case study indicated that the students developed a relatively high degree of planning competence with respect to the four major objectives defined for the course. Group performance as exhibited in the learning unit was superior to individual performance as indicated in the case studies. In comparing results for the two quarters, mean ratings were higher for the second quarter, but not significantly so. A wide range in planning competence was found.

Performance behavior.—Group meetings provided opportunity for observing student behavior in meeting actual problems involved in organizing the learning unit. Graduate students and teaching assistants in the School of Education observed the behavior of the students. The same criteria formulated for rating the written group projects were employed by the observers in evaluating the competence of each group's behavior. Individual members of each group were rated by the observers on a 4, 3, 2, 1 scale using as criteria six operational definitions of the fifth major objective, "Develops professional behavior." Each observer recorded the quality of individual student participation for each meeting in the following manner: a plus mark to indicate a positive contribution, a minus to indicate a delaying contribution, a question mark to indicate a question being asked, and a zero for an off-the-track comment. In addition, anecdotal information was recorded concerning individual and total group

TABLE 4.—RATING OF SEVENTY-FIVE OBSERVATIONAL CASE STUDIES, WINTER QUARTER

Criteria for Rating	Frequency of Rating				Mean Rating
	4	3	2	1	
1) Evaluates the effect of the home on the behavior of the child.	41	26	6	2	3.41
2) Collects and records information concerning the child in objective anecdotes.	45	17	9	4	3.37
3) Demonstrates understanding of how growth factors influence the behavior of a particular child.	36	22	14	3	3.21
4) Demonstrates understanding of how school adjustment influences behavior of the child.	31	24	15	5	3.08
5) Plans practical ways of helping the child.*	16	15	6	4	3.05
6) Demonstrates understanding of how the teacher may use guidance files and anecdotal records to evaluate the adjustment of the child.	26	25	16	8	2.92
7) Recognizes how the teacher may help the child achieve satisfactory relationships with age mates.	26	22	20	7	2.89
8) Demonstrates understanding of interests and abilities which influence the behavior of the child.	23	29	13	10	2.86
9) Checks hypotheses against an organized framework of psychological principles.	24	22	18	11	2.79
10) Forms a series of hypotheses to account for a particular pattern of behavior.	23	20	20	12	2.72

* Thirty-four were not rated on this criterion because their data did not warrant a plan of remedial action.

behavior. During the autumn quarter ten observers participated in the study; in the winter quarter there were six.

The range of mean ratings assigned by ten group observers for the ten criteria listed in Table 3 was from 2.80 to 3.70, autumn quarter; in the winter quarter from 2.83 to 3.67. For the autumn quarter the number of students rated 4, 3, 2, 1 was 39, 37, 19 and 8 respectively; in the winter quarter, 19, 19, 7, and 0, respectively.

A summary of the participation record and a report of descriptive anecdotes in chronological order for two students, autumn quarter, illustrates the data on which individual ratings were based: Student A, assigned a 4 rating; this student was the group recorder once; the group chairman, three times.

Participation record summarized:

From nine to twenty-three positive contributions per meeting.
From two to five questions asked per meeting.
No off-track comments.
No delaying comments.

Descriptive statements:

Always conscious of what group is doing. Good at directing; volunteered an assignment at end of meeting. (First meeting.)
Read report to group; a positive contribution.
Showed excellent leadership:
  "Let's do something more constructive."
  "We're to consider the psychology; how we're going to put the material across."
  Suggested assignment of definite tasks, extra meetings.
  Went after film catalog.
  Direct, practical, good organizer.
Has taken over chairmanship, is completely conscious of whole group situation, of each individual.
Organizes, allocates responsibilities, checks for performance.

Student D, assigned a 1 rating; this student served as recorder once; never as chairman.

Participation record summarized:

From three to eight positive contributions per meeting.
From one to three questions per meeting.
From one to four off-track comments per meeting.
From one to five delaying contributions per meeting.
Absent from one meeting; positive contributions decreased from beginning to end of quarter.

Descriptive comments:

> Takes project lightly.
> Amused over situation.
> Doesn't listen to others.
> Makes very few positive contributions
> Appears uninterested.

The first part of one observer's summary of group progress *illustrates change from beginning to end of quarter and indicates* quality of performance:

"At the first meeting three or four students did most of the talking. The others listened or hesitantly submitted questions. The time was spent in setting up a learning situation for the study. The next meeting was characterized by greater participation by all members; they were still not discussing learning but were emphasizing the desired learning situations. By the third meeting, most of the members realized they had accomplished little in regard to understanding and explaining the learning process. This understanding could be attributed to numerous causes such as outside reading, subgroup discussions, possibly discussion with other groups, and casual conversation with each other. This awareness was noticeably present during the third meeting, along with a determination to do something about it. This insight added vigor toward solution of their problem. Small subgroups were set up for gathering information and organizing it around the earlier learning situation or framework. The chairman of each group reported on his section of the work, made the necessary revisions suggested by the group and finally inserted his contribution in with the remaining body of material making up the Group 2, Section 2, Learning Report."

The data collected by the observers indicated that the students demonstrated satisfactory competence in meeting actual problems. Other data obtained in individual conferences and in recorders' reports submitted after each group meeting substantiated this conclusion. A range of competence was found among the students in meeting actual situations and in exercising professional behaviors.

Student opinion and *self-evaluation.*—Conclusions *reported* thus far have been based upon data obtained from objective-type tests, rating scales, and anecdotal information. In the questionnaire and self-evaluation form, students expressed their opinions concerning the course and estimated their own competence. These instruments were developed and administered to summarize in quantitative form (1) the effectiveness of motivation, (2) the degree of student acceptance of the objectives, (3) student opinion of course experiences, and (4) student self-rating.

The questionnaire consisted of two parts, the first containing fifteen questions concerning course activities and objectives, the second containing six questions relating to weaknesses and strengths of the course. In Part I, all students expressed the opinion that the course had been valuable in increasing their competence as teachers and in developing their understanding of the rôle of the teacher as an organizer of learning experiences. There were no negative responses to these items. Ninety-seven per cent felt that the course had increased their understanding of how children learn and of the rôle of the teacher in counseling and guidance. Ninety-four per cent felt that the course had made other educational experiences more meaningful. Ninety-two per cent indicated that the course had increased their understanding of child growth and development. Concerning specific course activities—lectures, classroom observations, learning unit, case study, and class discussions—six to eighteen per cent expressed the opinion that these had not contributed to achieving the purpose intended. Ninety-six per cent indicated that the course would not have been more valuable as a 'straight' lecture course. Understanding of psychological theory applied to the classroom was increased for ninety-four per cent. For twenty and fifteen per cent, respectively, the course did not open any new fields of interest nor was the course organization meaningful.

In Part II of the questionnaire, students listed all major activities as strengths of the course; all major activities were also listed at least once as weaknesses of the course. For improving the course, students listed more time for developing the various activities and also suggested increasing the unit credit from four to eight quarter-hours. Suggestions were made for improving all aspects of the course including selection of text materials. The data in Part II supported that obtained in Part I; the majority accepted the activities as valuable in achieving the objectives; for a minority one or more activities were not considered of value.

TABLE 5.—STUDENT RATING OF OWN COMPETENCE, WINTER QUARTER

Criteria for Rating	Frequency of Rating				Mean Rating
	4	3	2	1	
1) Evaluates the effect of the home on behavior of the child.	37	23	6	0	3.47
2) Understands use of guidance files and anecdotal records to evaluate adjustment of child.	33	25	8	0	3.38
3) Plans learning experiences which utilize present interest to focus attention on subject matter.	34	25	6	2	3.36
4) Collects and records information concerning pupil behavior in objective anecdotes.	31	27	8	0	3.35
5) Understands how growth factors influence the behavior of a particular child.	30	25	10	1	3.27
6) Plans learning experiences in which motivation is developed through positive interest.	22	30	14	0	3.12
7) Plans learning experiences which provide opportunity for successful participation of all pupils.	25	25	14	2	3.11
8) Plans practical ways for helping the child.	21	31	12	1	3.11
9) Understands the implications of social class training for the teacher.	26	25	10	5	3.09
10) Plans a variety of experiences to motivate a particular learning.	21	29	14	2	3.05
11) Understands conditions necessary for maintaining group morale in the classroom.	17	34	13	2	3.00
12) Plans experiences which take into account factors of readiness.	16	35	15	1	2.98
13) Plans experiences which provide an overview of the whole problem and its related parts.	23	23	13	6	2.97

TABLE 5.—(*Continued*)

Criteria for Rating	Frequency of Rating				Mean Rating
	4	3	2	1	
14) Forms a series of hypotheses to account for a particular pattern of behavior.	16	35	12	3	2.97
15) Recognizes how the teacher may help the child achieve satisfactory relationships with age mates.	17	31	13	5	2.91
16) Plans learning experiences which satisfy pupil needs.	15	30	21	1	2.88
17) Understands how the teacher may use standardized test results to evaluate the status of an individual child.	18	25	20	3	2.88
18) Participates in defining instructional objectives and in curriculum planning.	18	26	17	6	2.84
19) Plans experiences which ensure transfer.	14	31	16	5	2.82
20) Plans experiences which meet the needs of heterogeneous abilities and interests.	10	35	18	4	2.76

On a self-evaluation form, students rated themselves on a 4, 3, 2, 1 scale using as criteria for rating twenty operational definitions of the four major objectives. Nine of these criteria were identical to those employed in evaluating the learning unit; seven were identical to those used in rating the observational case studies. The results of this rating are summarized in Table 5.

Students in appraising own competence demonstrated a range both with respect to a particular criterion and different criteria. The general pattern was similar to that obtained in other evaluative instruments. On the criteria identical to those for the group learning unit, mean ratings were lower in the self-evaluation than for the group ratings. On criteria identical to those for the case study, the range of mean ratings on the self-appraisal was from 2.88 to 3.47; the mean ratings assigned to the written studies ranged from 2.72 to 3.41.

SUMMARY

A course in educational psychology for teachers was organized with objectives defined as teaching behaviors which behaviors were drawn from "Factors in Teaching Competence." Course experiences were organized to achieve the objectives, and evaluation procedures were devised to determine the extent to which the teacher competences were developed. The data collected and analyzed for two consecutive quarters with two groups of students indicate that:

1) The objectives as defined and delimited for the last quarter were psychologically valid.

2) The course experiences as organized and presented were effective in developing teaching competences as defined in this study.

3) The over-all strategy was effective in achieving the objectives.

4) At three levels—theoretical, planning, and performance—the majority of students exhibited a high level of competence. A small minority did not demonstrate growth or satisfactory competence at any level and were recommended as doubtful candidates for the teaching profession. Because of the paucity of data relative to pre-service standards, the conclusions in terms of 'high' and 'unsatisfactory' were necessarily tentative. It is recommended that the students be followed in their student-teaching and later teaching positions. The final criterion upon which the effectiveness of this or any professional course can be determined must be the quality of pupil growth developed in the teacher's classroom.

A STUDY OF TEACHING POTENTIALITIES*

J. DAVID O'DEA

School of Education, Oregon State College, Corvallis

This study is in the general field of the measurement of a set of characteristics objectively regarded as symptomatic of an aptitude for the teaching profession.

First will be presented the importance of this and similar studies. Secondly, a brief treatment of the early testing movement in the selection of teacher candidates will be given. Finally, a feasible procedure will be offered which will assist in the selection of teacher candidates.

IMPORTANCE OF STUDY

The research of Dr. Ray C. Maul[9] in the field of supply and demand of teachers has pointed to a need for additional studies. In reference to his educational research, Dr. Maul writes, "It is designed to help the public gain a fuller understanding of a problem in which the public has a vital stake, namely the improvement of schools in which the children of today must be prepared to meet the unpredictable problems of tomorrow."

To relieve this vexing problem the teaching profession must labor in the partnership with the farmers, ministers, merchants, and others engaged in helping children to become better adults. Obviously, the people needed to meet the present demand should possess qualities or traits of competent teachers.

RELATED STUDIES

The professional schools for training teachers have limited means for ascertaining in advance whether a candidate has or can acquire those qualities possessed by successful teachers. The measurement of teaching efficiency has engaged the attention of research workers for several years. Meriam,[8] in 1906, is credited as being the first who attempted to use objective measures of qualities related to teaching efficiency. This first attempt in removing the measurement of teaching efficiency and aptitude from the field of opinion to that of objective measurement was followed by numerous attempts. Because of one or many reasons these attempts had only limited success.

---

* For their constructive criticism, appreciation is extended to Ralph Evans, Head of the Department of Education, Fresno State College and to W. M. Ehrsam, Director of Guidance, Wichita University.

Ruediger and Strayer[11] in 1910 constructed a device to measure characteristics of teachers as related to efficiency in teaching. This attempt did not prove highly successful. In 1912, Elliot and Boyce[3,4] devised score cards for the measurement of teaching efficiency. The score cards for rating teachers were marked by the teacher's supervisors. These authors concluded that the reports were too high, due to a halo effect of the ratings on the score cards.

TABLE I.—AVERAGE CORRELATION OF SUPERVISORS' RANKINGS AND THE LISTED MEASURED CHARACTERISTICS

Quality or Trait	r
Scholarship	+.000
Intelligence	+.560
Professional Study in Service	+.393
Salary	+.352
Total Teaching Experience	+.172
Scores on a Professional Test	+.147

Knight[6] made an objective study on general teaching ability. His method was to obtain ratings of teachers by both supervisors and the teachers' associates in teaching upon general professional ability. Each judge ranked the teachers whom he knew in order of merit. These rankings were then transmuted into numerical units representing the composite judgment of each group of judges. Against this criterion he correlated the measurements made upon a number of traits which could be measured objectively. The average correlations found for high school teachers are shown in Table I.

Knight applied the same method used for high school teachers to the study of the elementary school teachers. Two of his results should be noted: first, a correlation of +.541 between teaching efficiency of elementary school teachers and a professional test; and second, a correlation of +.108 between teaching efficiency and scores on an intelligence test.[9] Knight concluded that a professional test is highly predictive of teaching efficiency for elementary school teachers, but he suggests that an intelligence test may be more desirable for prediction on the secondary level.*

In 1923, Somers[12] obtained ratings of teachers in their first year of experience. He used a score card and correlated against this criterion scores in intelligence tests, school marks, participation in extra-curricular activities and personality estimates. The most significant correlation co-efficients are shown in Table II.

TABLE II.—CORRELATION OF TEACHERS' RATING IN THEIR FIRST YEAR EXPERIENCE AND LISTED CRITERIA

Factors Used	r
Normal School Success	+.73
Two Years' Marks in School	+.71
Teaching in Training School	+.70
Personality Estimates	+.62
First Semester Normal School Work	+.60
Intelligence Tests	+.43
Extracurricular Activities	+.41
Class "Cuts" in Normal Schools	+.53

A study by Whitney[13] followed a similar technique used by Somers. The correlations between the criteria and teaching success as measured by supervisors are as follows:

TABLE III.—CORRELATION OF TEACHERS' RATING IN THEIR FIRST YEAR EXPERIENCE AND THE LISTED CRITERIA

Factors Used	r
Intelligence	+.025
Secondary School Marks	+.091
Marks in Academic Subjects in Normal School	+.073
Marks in Professional Subjects in Normal School	+.143
Practice Teaching in Normal School	+.238
Physique	+.124

Knight, Ruch, Telford, and Bathurst[7] have published a set of professional tests for elementary school teachers. These aptitude tests were the first for which a measure of validity was obtained. The authors obtained a correlation co-efficient of +.378 using the test scores and the merit of ranking of the teachers by supervisors.

---

* Dr. C. W. Boardman has written an excellent condensation of Knight's study. His research, reference,[3] was particularly helpful to this study.

Corresponding to Knight's study is the George Washington University Series, prepared by F. A. Moss, T. Hunt, and F. C. Wallace in the coöperation with various educators.[10]

The progress in the field of teacher aptitude tests is illustrated by Bingham.[3]

> "The professional schools for training teachers have as yet found no sure means of ascertaining in advance whether a candidate has or can acquire the traits described in the foregoing paragraphs, although marked defects and shortcomings are usually recognizable during an interview with an applicant for admission. His intellectual grasp and mental alertness can be fairly well estimated by references to his previous school record and his performance in tests of scholastic aptitude, vocabulary, English usage, mathematics, and other academic subjects, supplemented when necessary by specially devised tests of intelligence such as the Coxe-Orleans Prognosis Test of Teaching Ability."

The Coxe-Orleans Test is a device to measure the ability to *learn the subject matter taught in school. The five parts of the* test measure the candidate's general information, knowledge of teaching methods and practices as acquired from his observation as a student, ability to learn the type of material included in professional reading matter, and ability to study and work out educational problems.

The co-efficients of correlation between test scores and achievement at the end of the first college year range between .53 and .84.

It is noted that, although the authors have demonstrated the value of this test for predicting school achievement, the extent to which it is prognostic of actual teaching ability has not been ascertained.

A recent reference pertaining to the measurement and prediction of teaching efficiency is a monograph by Barr[1] and others. The monograph, "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigations," supplies data on the validity and reliability of some 72 tests, rating scales and inventories frequently used in this area.

SELECTION OF THE MEASURED CHARACTERISTICS

A test to measure the characteristics listed in Table IV was constructed under the direction of Dr. H. E. Schrammel, Director of the Bureau of Measurements, Kansas State Teachers College, Emporia.

Studies indicate that personality and intelligence are highly significant factors for measuring teaching aptitudes. Knight found these factors significant to teaching efficiency in high school. These two basic traits were not measured in this study.

Table IV lists the set of characteristics that were selected for measurements.

TABLE IV.—CHARACTERISTICS MEASURED IN THIS STUDY

Test	Form	No. of Questions
1. Interest	Multiple Response	15
2. General Knowledge	" "	45
3. (a) Procedures in Teaching	" "	20
(b) Teaching Information & Judgment	True—False	67
4. Prof. Reading Comp.	" — "	25
Total		172

TRYOUT OF TEST

The control group in this study consisted of teachers. The teachers coöperating in the study were probably a select group of professionally minded teachers who would tend to rank above the average of the school teaching population. They were elementary, high school, and college teachers all of whom were members of the faculty at the Kansas State Teachers College, Emporia.

The second group consisted of college students. This group was broken down into various categories as indicated in Table V.

In summary, nearly three hundred tests were given. The time required to take the test averaged about fifty minutes.

SUMMARY OF STATISTICAL DATA

The majority of the statistical measures were computed from the test results obtained from the administration of the test at the Kansas State Teachers College. The basic data are shown in Table V. Additional data computed from the basic data are listed in Table VI.

TABLE V.—COMPARISON OF VARIOUS STATISTICAL MEASURES

	Control	All of the Experimental Group	Non-teaching Group	Experimental Group, going to Teach
Median	147.1	120.0	120.3	127.3
Mean	146.9	125.0	119	127.2
Standard Deviation	5.9	13.7	13.5	12.2
Sigma Mean	1.07	1.00	1.82	1.05
Number of Cases	30	189	57	132

Read Table thus: The median for the control group was 147.1; for the entire group, 120.0.

TABLE VI.—COMPARISON OF FURTHER STATISTICAL MEASURES

	Difference	Sigma Difference	$\frac{\text{Difference}}{\text{Sigma Difference}}$ (Critical Ratio)	Number of Chances in 100 of True Difference
All of the Experimental Group	21.9	1.46	15.0	100
Non-teaching Group	27.6	2.10	13.1	100
Experimental Group Going To Teach	19.7	1.50	13.1	100

Read Table thus: The difference between the control group true mean and all the experimental group true mean was 21.9.

Comparison of the mean scores shows that the control group made a higher score than did either of the experimental divisions. The mean score for the control group was 146.9 and that of the total experimental group, 125.0.

The data listed in Table VI deal with the reliability of the difference between the mean scores of the control group and the experimental groups. It will be noted that because these differences are large and the sigma differences are small, the data yield a statistically significant ratio. In other words, there are 100 chances in 100 that the differences possess statistical significance.

By correlating the odd and even scores of the test a reliability co-efficient of .77 was obtained. The P. E. was ±.009. By correlating the combined scores of Parts I and II with Parts III and IV, Table IV, a reliability co-efficient of .88 was obtained. The P. E. was ±.007.

SUMMARY AND CONCLUSION

This study attempted to isolate a set of measurable characteristics indicative of teaching success. The test used to measure the characteristics indicates that the control group possessed the measured traits to a higher degree than those of the other group. In general, those preparing for teaching score higher than those preparing for some other vocation. This finding is vitally significant to the study.

It is found that a good teacher need not score high in one singled out characteristic. However, a good teacher should possess in total a high degree of knowledge, intellect, interest, professional reading comprehension, and teaching procedures.

Personality has long been held to be a large factor in teaching success. Past studies have indicated that such traits as self-control and patience are essentials to teaching success.

It is the opinion of the author that the following procedures will provide some indications necessary in attempting to evaluate those traits indicative of teaching potentialities. These factors are: (a) the student's attitude regarding further education and his like or dislike of previous educational experiences. This involves the student's previous school success. (b) The student's personality should be investigated. This includes mental ability, emotional pattern, and the work habits of the student. (c) Finally, careful consideration of the measured characteristics as listed in Table IV is statistically significant in the selection of potential teacher candidates.

REFERENCES

1. Archer, C. P., Barr, A. S., Nelson, M. J., Raths, L. E. and Reitz, W., "The Measurement and Prediction of Teaching Effi-

ciency: A Summary of Investigations," Dembar Publications, Inc., 1948.

2. Bingham, Walter Van Dyke, *Aptitudes and Aptitude Testing.* Harper and Brothers Publishers. New York and London. Copyright 1937, pp. 199.

3. Boardman, C. W., *Professional Tests as Measures of Teaching Efficiency in High School*, Teachers College, Columbia University, No. 327, 1928.

4. Boyce, A. C., "Methods for Measuring Teachers' Efficiency," *Fourteenth Yearbook*, National Society for the Study of Education, Part II, 1915.

5. Elliott, C. E., "A Provisional Plan for the Measure of Merit in Teaching," Report, Wisconsin State Board of Education, 1912.

6. Knight, F. B., *Qualities Related to Success in Teaching*, Teachers College, Columbia University, Contributions to Education, No. 122, 1922.

7. Knight, F. M., Ruch, G. M., Telford, F., and Bathurst, J. E., "Professional Tests for Elementary School Teachers," Public Personnel Studies, Vol. IV, No. 10, pp. 274–298, Oct., 1927.

8. Meriam, L. L., *Normal School Education and Efficiency in Teaching*, Teachers College, Columbia University, Contributions to Education, No. I, 1906.

9. Maul, R. C., "Teacher Supply and Demand in the United States," 1949 National Teacher Supply and Demand Study, Sponsored by the National Educational Association of the United States.

10. Moss, F. A., Hunt, T., and Wallace, F. C., "Teaching Aptitude Test," George Washington University Series, Copyright 1927.

11. Ruediger, W. C., and Strayer, G. D., "The Quality of Merits in Teachers," Journal of Educational Psychology, Vol. I, pp. 272–278, May, 1910.

12. Somers, G. T., *Pedagogical Prognosis: Predicting the Success of Prospective Teachers*, Teachers College, Columbia University, Contributions to Education, No. 140, 1923.

13. Whitney, F. L., "Prediction of Teaching Success," *Journal of Educational Research*, Monograph No. 6, 1924, Public School Publishing Company, Bloomington, Illinois.

TECHNIQUE FOR EVALUATING THE ABILITY OF TEACHERS TO APPLY PRINCIPLES CONCERNED WITH THE DEVELOPMENTAL NEEDS OF ADOLESCENT GIRLS*

SARA ANN BROWN

West Virginia University

The purpose of this study was to develop a technique for evaluating the ability of secondary-school teachers to apply the principles concerned with the developmental needs of adolescent girls. The first hypothesis was that pencil-and-paper tests would be developed which would measure the ability of teachers to recognize and apply principles concerned with human growth and development of adolescents. Two tests were developed to measure this ability: Test A, Recognition of Principles; and Test B, The Case of Jacqueline Croner, essay and short-answer forms. The second hypothesis was that one case study would be sufficient to measure this ability.

The assumption was that it is important for all teachers to be able to apply principles concerned with the developmental needs of children. The experimental group in the study was composed of home economics teachers at the secondary level, on the further assumption that instruments which were valid and reliable for that group would be equally useful for all secondary-school teachers.

Test A, Recognition of Principles, was made up of sixty-two principles which related to the needs of adolescents in three major aspects of development: physiological, social, and integrative. A jury of four judges assisted in the selection of the principles and agreed unanimously on the key for the short-answer test.

Test A was revised after having been administered to a preliminary group of thirty-five home economics teachers in West Virginia to discover any lack of clarity in vocabulary and statement, any difficulties in administration, and to determine the length of time required to give the test.

Test B, The Case of Jacqueline Croner, was designed to measure the ability which teachers have to apply principles in interpreting data concerned with the developmental needs of adolescents, and to select educational means for helping pupils with their problems. The test was a description of an adolescent based on data from a cumulative record of a high-school girl. A short-answer form which used responses furnished by a preliminary group of home economics teachers was set up. These responses were in the form of statements and were evaluated for their clarity and conciseness. The test was administered to the same preliminary group as the one which reacted to Test A. Revisions based on the results from the preliminary group were made. The revised test which contained one hundred ten items was administered to a jury of six judges, and from their responses the key for scoring was determined.

---

* Abstract of Doctoral thesis submitted to Graduate Faculty, Iowa State College, 1949.

The trial group to whom the tests were administered was made up of eighty-eight home economics teachers in West Virginia who had signified a willingness to participate. Several factors were considered in selecting these teachers: marital status, recency of graduate study, length of teaching experience, and the feasibility of visits by the writer.

Test B, The Case of Jacqueline Croner, essay form, was sent to each of the participants. This was followed by Test B, short-answer form; and Test A, Recognition of Principles.

The reliability of Test A was determined by obtaining the correlation coefficient between scores on split halves. A Pearsonian coefficient of correlation was obtained, .7305, and corrected for double length by the Spearman-Brown modified formula. This correction yielded a coefficient of reliability of .848 ± .030. Since testmakers cite a relationship of .90 as very high this reliability is considered very satisfactory.

The reliability of Test B, short-answer form, was determined by the same method and when the correlation of .8379 was corrected for double length, it was found to be .913 ± .018, a result which would indicate a very dependable reliability.

Two methods for establishing validity of the instrument were employed. One was the correlating of scores on Test A and on Test B, essay form, as well as the scores on Test A and Test B, short-answer form. The coefficients obtained were:

Test A and Test B, essay	.9554 ± .009
Test A and Test B, short-answer	.9758 ± .005
Test B, short-answer, and Test B, essay	.9363 ± .013

In order to have greater confidence in the results, coöperation from another group of ninety home economics teachers in West Virginia, who had not participated in the first group, was requested. The procedure except for visits and interviews was duplicated and the coefficients of correlation were again obtained.

The results for the second group were:

Test A and Test B, essay	.9506 ± .010
Test A and Test B, short-answer	.9557 ± .009
Test B, short-answer and Test B, essay	.9247 ± .015

These results would seem to indicate that the high correlations of the first group were not due to sampling, since the differences in the correlations are slight.

A second method of validation which was used, was to compute coefficients of correlation between the performance on the instruments with a criterion assumed to be valid. In this study that criterion was the scores obtained on interviews which the writer had with the teacher, the principal and/or the superintendent, and the score on a one-day observation of the teacher, designated as an interview-observation score. Records of these interviews and observations with the group of eighty-eight teachers were kept by the writer, and these were scored by a jury of five judges. When coefficients of correlation were computed between the scores on the tests and the interview-observation scores, the correlations were:

Test A and interview-observation scores	.8748
Test B, essay form and interview-observation scores	.9013
Test B, short-answer and interview-observation scores	.9113

Since Test B was designed to measure the ability of the teacher to diagnose problems and select remedial measures, coefficients of correlation to determine whether one section of the test was more valid than the other were computed. The correlations between the diagnostic sections on the two tests—Test B, essay and Test B, short-answer form—was .7880 ± .040. The correlation between the scores on the two remedial sections was .4327 ± .087. Also computed were coefficients of correlation between the scores on the remedial section of the short-answer form of Test B and the interview-observation scores; the result was a coefficient correlation of .7866 ± .041.

The items on both Test A and Test B were analyzed to determine internal consistency and to discover items which should be discarded or revised because they were not discriminating. The eighty-eight papers were divided into two groups: the upper group was composed of the fifty per cent who received the highest scores; the lower group of the fifty per cent who received the lowest scores. The percentage of correct responses for each item was determined for each group. The items were then analyzed to determine (1) those for which each group had the same percentage of correct responses; (2) those for which the lower group had a higher percentage of correct responses; and (3) those for which the difference in percentage between the two groups was less than 10. All except thirteen of the sixty-two items in Test A, and thirty items of the one hundred ten in Test B, were satisfactorily discriminating.

In a teacher-education program it would be valuable if devices which could be administered easily were available, so that a prediction of what a teacher would do relative to the application of principles when he is faced with a real-life situation could be made. In the present study paper-and-pencil tests which have high correlations of validity and which can be used as the independent variables in computing regression equations have been developed. For the criterion, the interview-observation score, which was used as the dependent variable in determining regression, the validity was assumed.

If a relationship exists between two distributions it becomes possible to predict values in one distribution from known values in the other. Regression lines are developed for the purpose of estimating values of one variable from known values of another variable within the limits of the data which are available.

In the following discussion of regression, the scores on the two tests and the interviews and observations will be designated by these symbols:

X_1 = Test A: Recognition of Principles Test
X_2 = Test B: The Case of Jacqueline Croner (short-answer form)
X_3 = Test B: The Case of Jacqueline Croner (essay form)
Y_1 = Interview-observation.

There were three main problems to be considered in analyzing the regression data. The first problem was: Can a regression equation which will predict teacher behavior in a real-life situation with reasonable accuracy be determined? The second problem was: How accurate would the prediction be if only the short answer tests, X_1 and X_2 were used? The third problem was: To what degree have all factors determining teacher behavior relative to applying principles concerned with adolescent development in a real-life situation been measured by the regression equation?

The interview-observation score was obtained from information relative to how the teacher applied principles concerned with the developmental needs of adolescent girls in a real-life situation. It was determined by the use of a scale which had a range of 1 to 15.

The first problem was to set up regression equations. The equations using all three variables is:

$$Y = -0.0096 - .2235X_1 + .2556X_3 + .2517X_2$$

The equation using two variables is:

$$Y = .0741X_1 + .2644X_2 - 5.9562$$

The equation using one variable is:

$$Y = 2140X_1 + 1.9382$$

To ascertain the usefulness of these equations it was necessary to determine the multiple coefficients obtained when all variables were used or progressively dropped; to determine the standard partial regression correlations holding various ones of the variables constant, and to make tests which indicate the significance of regression coefficients.

The multiple correlation coefficient obtained between all the estimated Y values and the actual values was .9366. This correlation for the three pencil-and-paper tests combine is sufficiently high to give reasonable confidence in prediction of the ability of teachers in this area.

The multiple R^2 is .8773. This correlation means that 87.73 per cent of the variability among teachers in their ability to apply principles in this area can be measured by the three pencil-and-paper tests.

If in Test B, X_3 is omitted from the multiple correlation, R^2 becomes .8394. Thus the difference between the multiple R^2 based on three independent variables and the multiple R^2 based on two independent variables is .0424. If one independent variable is dropped from the equation an additional 4.24 per cent of the variability remains unaccounted for by regression.

The standard partial regression coefficient specifies changes in Y, independent of the changes in the other independent variables and is the best measure of which variable is the most important in predicting Y. The b' for each of the three tests is:

$$X_1 = .9139$$
$$X_2 = 1.1492$$
$$X_3 = .6984$$

It is obvious that X_2 is the most important test in predicting Y, since the standard partial coefficient for it is the highest of the three tests.

Standard partial regression coefficients may also be used to judge which of the independent variables are most important in estimating Y. If X_2 is held constant and X_1 is compared with Y, the correlation is low—.0160. If X_3 is held constant and X_1 is compared with Y the correlation is low—.1071.

If X_1 is held constant the correlations between X_2 and Y, and between X_3 and Y are:

$$r_{Y2\cdot 1} = .5449$$
$$r_{Y3\cdot 1} = .4577$$

When X_3 is held constant the correlations between X_1 and Y, between X_2 and Y, and between X_1 and X_2 are:

$$r_{Y1\cdot 3} = .1071$$
$$r_{Y2\cdot 3} = .4431$$
$$r_{12\cdot 3} = .7840$$

When X_1 is held constant, the correlation between X_3 and X_2 is:

$$r_{32\cdot 1} = .0619$$

When each independent variable is compared with Y, the other two variables being held constant, the following correlations are obtained.

$$r_{Y1\cdot 23} = .4318$$
$$r_{Y2\cdot 13} = .5818$$
$$r_{Y3\cdot 12} = .5495$$

Since all t tests coefficients—4.385, 5.385 and 6.559—are highly significant, each of the independent variables may be assumed to be related to the dependent variables in a linear manner. All three also contribute to the total regression, and for the most accurate prediction relative to the ability of teachers to apply principles in this area, all three variables should be measured.

The extremely high value of multiple R^2—.8773—is evident of the success with which individual predictions may be made when the three independent variables selected in this study are used.

Throughout the statistical analysis, X_3, the essay form of Test B, The Case of Jacqueline Croner, has proved to be the least valuable of the variables to use for prediction purposes. It has the smallest standard partial regression coefficient, the smallest partial correlation with the Y variable, and the smallest t value. It seems, therefore, that it is justifiable to drop it from the regression equation, and lose a very small part of the total predictable information in favor of more easily administered and time-saving tests.

The recommended regression equation obtained is:

$$Y = -.0741X_1 + .2644X_2 - 5.9562$$

This Y variable may be predicted for individuals by measuring the X_1 and X_2 variables, a task which would take considerable time, since approximately one and three-fourths hours are required for each individual to read the two tests, and additional time is necessary to score it.

It can be reasonably assumed that the ability of individuals to apply principles can be measured by the use of X_1 and X_2. These tests should make it unnecessary for teacher-educators to expend time, energy and money in visiting widely distributed schools for the purpose of observing teachers in their classrooms and, during extra-curricular activities, of interviewing them and consulting their administrators.

THE USE OF LITERAL GRADES

TRUMAN L. KELLEY

Santa Barbara, California

A system of rating, or classifying, students by means of letters A, B, C, etc., or E (excellent), G (good), etc., is called a literal grading system. It commonly gives an order of merit of groups, with no very precise quantitative relationships between the literal grades.

The distaste of the statistician for literal grades is fully justified on two counts: Generally but few classes are employed so that there is a large grouping error and as the literal grades provide an ordered series only the quantitative relationships between the classes is unknown. We here propose literal grades which shall have these shortcomings to as small a degree as possible. A solution is possible if the form of distribution of the trait being graded is known.

We shall assume a normal distribution of talent, arguing that the abilities, pressures and interests that bring students into classes tend to produce a homogeneous group, that is, a group that does have a modal ability-pressure-interest characteristic and frequency falling off from this mode in both directions in a somewhat normal manner.

For notation we use the following: The class given the literal grade A is the highest ranking class, the next highest class has literal grade B, etc. The A class is the tail of a normal distribution. Assuming a unit normal distribution its mean is $\bar{A}$, its tempered class index (described later) is A, its upper limit is infinity and its lower limit which is at the boundary between classes A and B is x_{AB}, its ordinate at the lower boundary is z_{AB}, the number of cases in the class is q_A, which is identical with q_{AB} the proportion to the right of the point x_{AB}, the variance of the measures in class A from the class mean $\bar{A}$ is V'_A, and finally the mean squared deviation of the measures in this class from the tempered class index, A, is here called the class variance and is labelled V_A. For class B we have notation

$\bar{B}$ = class mean

B = tempered class index

x_{AB} = upper class limit
x_{BC} = lower class limit
z_{AB} = ordinate at the upper limit
z_{BC} = ordinate at the lower limit
$q_B = q_{BC} - q_{AB}$ = proportion of cases in the class
V'_B = variance from $\bar{B}$
V_B = variance from B

Of course we use similar notation for all other classes.

Except for a two-class system, A and B, or a three-class system A, B, and C, it is impossible to have class means such that all differences between neighboring means are equal. However it is almost universal to assume that the A pupils are as superior to the B pupils as the B pupils are superior to the C pupils, etc. Since $\bar{A}$, $\bar{B}$, $\bar{C}$, $\bar{D}$, etc. do not show equal differences we modify, or temper, the scale employing indexes A, B, C, D, etc. upon which has been imposed the condition $A - B = B - C = C - D$, etc. This is the first imposition that we make. The second is that the class limits shall be so chosen that the variance error of estimate due to grouping

$$V_{est} = q_A V_A + q_B V_B + q_C V_C + \text{etc.}$$

shall be minimal.

Even when those in a class, say B, are correctly assigned to the class there are errors when all are given the same mark B, for in ability they range all the way from x_{BC} to x_{AB}. The variance of the measures from B is the variance error due to grouping for those graded B. This is not the total variance error of estimate, which is equal to the variance due to grouping plus the variance due to misjudgment which has placed individuals in wrong classes. We here only attempt to minimize the error due to grouping.

We shall not linger over the two category case for it is surely axiomatic that to minimize the grouping error the proportion in the A class should be .50 and the proportion in the B class also be .50.

It is obvious in the three category case that the proportion in the A class must be the same as that in the C class, so class means $\bar{A}$, $\bar{B}$, and $\bar{C}$ must be equally spaced and thus constitute the tempered class indexes A, B, and C. Utilizing formulas [8:26] and [8:29] of Kelley's *Fundamentals of Statistics*, 1947, we obtain

$$V_A = V_C = 1 + \frac{x_{AB}z_{AB}}{q_A} - \left(\frac{z_{AB}}{q_A}\right)^2$$

$$V_B = 1 - \frac{2x_{AB}z_{AB}}{1 - 2q_{AB}}$$

Weighting these three variances by the respective proportions involved we obtain the function which we desire to minimize. The minimum occurs when $z_{AB} = 2q_Ax_{AB}$. It is interesting to note that this is the same condition as that which gives the necessary upper and lower tail portions of a normal distribution to maximize the critical ratio of the difference between means of tail portions over its standard error (see Kelley, *Fundamentals of Statistics*, Ch. VIII, Sec. 8). At this point $q_A = .27027$ and $x_{AB} = .61200$. Table I herewith is given to facilitate comparison with later tables for 4, 5, and 6 category classifications.

TABLE I.—THREE OPTIMAL CLASSES FOR A BASIC UNIT NORMAL DISTRIBUTION (σ = 1.00)

Class	Class Limits		Proportions in Classes	Actual Class Means	Tempered Class Indexes	Intra-class Variance from Class Means
	Lower $x =$	Upper $x =$				
A	.61200	∞	.27027	1.2240004	1.2240004	.25090
B	−.61200	.61200	.45940	.0000000	.0000000	.11873
C	−∞	−.61200	.27027	−1.2240004	−1.2240004	.25090
	Weighted intra-class variance					.19017
	Standard error of measurement due to grouping					.43609

The number of decimal places in the tables which follow is, for most purposes, excessive. They have been reported to this large number of places for two reasons. In the case of Table II, and the lower limit of class A, it was necessary to compute a large number of decimal places to establish, indubitably, that the class limit was not exactly one standard deviation from the mean. In all tables, to secure proportions correct to three or four places many additional places were necessary in the class limits. Any one desiring to check the writer's work, or to proceed further with this sort of determination, will need the eight-decimal place values of x as given.

The determination of the proportions entering into the optimal four-class system was accomplished by cut-and-try methods. The outcome is given in Table II. This is easy to remember because x_{AB} is almost exactly one standard deviation.

TABLE II.—FOUR OPTIMAL CLASSES FOR A BASIC UNIT NORMAL DISTRIBUTION (σ = 1.00)

Class	Class limits		Proportions in Classes	Actual Class Means	Tempered Class Indexes	Intra-Class Variance from Tempered Class Indexes
	Lower x =	Upper x =				
A	.99960851	∞	.15875	1.52482177	1.46400843	.202842
B	.00000000	.99960851	.34125	.45971231	.48800279	.080393
C	−.99960851	.00000000	.34125	−.45971231	−.48800279	.080393
D	−∞	−.99960851	.15875	−1.52482177	−1.46400843	.202842
	Weighted intra-class variance					.119271
	Standard error of measurement due to grouping					.345356

Using the same laborious methods solutions have been obtained for five-class and six-class systems. The results are shown in Tables III and IV.

Our general conclusion is that there can be an improvement in literal grading systems if the proportions in the classes are well chosen. The well chosen proportions, if the trait is normally distributed, are given for three-, four-, five- and six-class systems in Tables I, II, III and IV.

TABLE III.—FIVE OPTIMAL CLASSES

Class	Class limits		Proportions in Classes	Actual Class Means	Tempered Class Indexes	Intra-Class Variances from Tempered Class Indexes
	Lower $x =$	Upper $x =$				
A	1.26575727	∞	.1028	1.74187247	1.66317989	.170801
B	.42450105	1.26575727	.2328	.79084081	.83158904	.057438
C	−.42450105	.42450105	.3288	.00000000	.00000000	.058037
D	−1.26575727	−.42450105	.2328	−.79084081	−.83158904	.057438
E	−∞	−1.26575727	.1028	−1.74187247	−1.66317989	.170801
	Weighted intra-class variance					.082385
	Standard error of measurement due to grouping					.287028

TABLE IV.—SIX OPTIMAL CLASSES

Class	Class limits		Proportions in Classes	Actual Class Means	Tempered Class Indexes	Intra-Class Variances from Tempered Class Indexes
	Lower $x =$	Upper $x =$				
A	1.40105027	∞	.072	1.90562007	1.80680381	.102593
B	.73884085	1.40105027	.158	1.05343005	1.08108228	.042380
C	.00000000	.73884085	.270	.35204148	.36136070	.044570
D	−.73884085	.00000000	.270	−.35204148	−.36136070	.044570
E	−1.40105027	−.73884085	.158	−1.05343005	−1.08108228	.042380
F	−∞	−1.40105027	.072	−1.90562007	−1.80680381	.102593
	Weighted intra-class variance					.060878
	Standard error of measurement due to grouping					.240735

MENTAL ABILITY RATINGS OF HONOR STUDENTS

LILLIAN G. PORTENIER

University of Wyoming

Do all honor students earn superior ratings on tests of mental ability? What is the range of mental ability scores for students upon whom a given honor has been conferred? How widely do mean mental test scores vary for students receiving different honors? To answer these and other questions, data were compiled for students who were the recipients of awards and/or who were elected to various professional and honorary fraternities at the University of Wyoming during the twenty year period from 1929 to 1949.

In common with most institutions of higher learning the University of Wyoming attempts to promote interest in, and to grant recognition for, outstanding work in the different lines of activity by election to various collegiate honorary societies, and by conferring awards, prizes and scholarships. Printed programs listing the honorary elections and awards are distributed at the 'Honors' assembly held annually during the latter part of May. Copies of the twenty programs from 1929–1949 furnished a complete list of the honorary elections and award recipients for the period.

Since the Ohio State University Psychological Examination has been administered to freshmen at the University annually, beginning with the fall quarter 1929, scores were available in the personnel files for most of the honor students. Successive forms of the O. S. U. Test from Form 14 to Form 23 were used during the twenty years. To secure comparable scores for students who had taken different forms of the test, percentile ratings based on freshmen norms were used. The mean and range of the percentile ratings for students receiving each of the honors listed are shown in Table I. A study of these data reveals many significant facts. First, there is a very wide range in percentile ratings for the recipients of most awards, in a few cases from the lowest to the highest decile. For the Phi Beta Kappa awards only were all the percentile ratings in the upper quartile. The mean percentile ratings for the recipients of awards from the honorary fraternities are higher than the mean ratings for

TABLE I.—RANGE AND MEAN OF PERCENTILE RATINGS ON THE O. S. U. TEST FOR STUDENTS RECEIVING PROFESSIONAL AND HONORARY FRATERNITY AWARDS AT THE UNIVERSITY OF WYOMING, 1929–1949*

Organization	Number	Range	Mean
Alpha Epsilon Delta			
Outstanding premedical freshman	8	60–99	87.4
Alpha Kappa Psi			
Outstanding junior (male) in Commerce	10	40–99	78.6
Alpha Zeta			
Agriculture freshman with highest scholastic ave.	14	9–93	59.1
American Society Mechanical Engineers			
Outstanding work in Student Branch A. S. M. E.	12	13–95	63.0
Gamma Sigma Epsilon			
Highest ranking first-year chemistry student	20	8–100	80.8
History Club			
Highest ranking freshman in history	18	37–100	84.4
Kappa Delta Pi			
Educ. fresh. and soph. with highest scholastic ave.	40	20–100	84.9
Phi Beta Kappa			
Graduating senior with highest scholastic ave.	16	76–100	90.3
Phi Kappa Phi			
Outstanding freshman, sophomore and junior	59	9–100	89.5
Phi Gamma Nu			
Woman major in Commerce with highest scholastic ave.	16	26–100	84.5
Phi Sigma Iota			
Outstanding freshman in French and in Spanish	36	28–100	84.2
Phi Upsilon Omicron			
Highest ranking freshman in Home Economics	20	61–99	82.4

TABLE I.—(*Continued*)

Organization	Number	Range	Mean
Pi Delta Epsilon			
Outstanding first year Branding Iron reporter	15	46–100	78.7
Psi Chi			
Undergraduate showing promise in psychology	9	71–98	88.6
Sigma Tau			
Outstanding freshman in Engineering	9	12–98	81.0
Theta Alpha Phi			
Outstanding actor and actress in the Univ. Theatre	40	29–99	76.2

* Many of the above fraternities have been installed at Wyoming since 1929, and some awards were not conferred annually. Most of the awards were made for at least ten years although O. S. U. scores were not available for all recipients.

recipients of awards from professional organizations. Phi Beta Kappa (90.3) has the highest mean rating followed closely by Phi Kappa Phi (89.5). The lowest mean ratings were found for Alpha Zeta (59.1) which confers its award on the freshman earning the highest scholastic average in the College of Agriculture, and for the American Society of Mechanical Engineers (63.0) which recognizes outstanding work in the Student Branch of the Society. The mean ratings for all other groups fall in the upper quartile, and with but two exceptions, Alpha Kappa Psi (78.6) and Pi Delta Epsilon (78.7), in the second highest decile.

Most of the O. S. U. tests were administered to the students during the last semester of their senior year in high school or as entering freshmen at the University. In many instances the students did not realize the significance of the tests, and consequently their scores may not be reliable measures of the abilities tested. This fact probably accounts for most of the low ratings listed under the percentile range. However, many studies indicate that the O. S. U. Test, as well as most other tests designed as tests of general intelligence for college students, measure the abilities required for courses in the Liberal Arts and Sciences more successfully than they measure the abilities stressed in Agriculture and Engineering courses. (See reference[6].) This fact is verified not only by the relatively low mean percentile ratings for the recipients of the Alpha Zeta and A. S. M. E., awards but likewise by the mean ratings for the students elected to the three major honor societies: Phi Beta Kappa, Phi Kappa Phi, and Sigma Xi, as disclosed in Table II.

TABLE II.—RANGE, MEAN, STANDARD DEVIATION, AND CRITICAL RATIO OF PERCENTILE RATINGS ON O. S. U. TEST FOR STUDENTS ELECTED TO HONORARY ORGANIZATIONS AT THE UNIVERSITY OF WYOMING, 1929–1949

Organization	Number	Range	Mean	SD	CR
Phi Beta Kappa Outstanding scholarship in Liberal Arts and Sciences	56	28–100	92.32	12.85	
					3.88
Phi Kappa Phi Outstanding scholarship in any college	247	25–100	84.28	18.05	
					4.00
Sigma Xi Special aptitude for scientific research in pure and applied science	78	10–100	71.73	25.85	
					6.06

The mean percentile rating for Phi Beta Kappa (92.32), which limits its membership to outstanding scholars with their major course work in the Liberal Arts and Sciences, is eight points above the mean for Phi Kappa Phi (84.28), which elects outstanding scholars from any college, and twenty points higher than for Sigma Xi (71.73), which bases its recognition on special aptitude for scientific research in both pure and applied science. At Wyoming many students are elected to Sigma Xi from the Colleges of Agriculture and Engineering. The differences between the means are significant for all three groups as indicated by critical ratios of 3.88, 4.00 and 6.06, respectively.

The variability in the distribution of the percentile ratings is smallest in the Phi Beta Kappa group, S. D. 12.85. Only five (nine per cent) of the fifty-six Phi Beta Kappa students had percentile ratings below 75 and only one student below 60. The variability is far more pronounced for the Phi Kappa Phi group as shown by a SD of 18.05, and even more so for the Sigma Xi group with a SD of 25.85. More than twenty-five per cent of the students elected to Phi Kappa Phi and fifty per cent of those elected to Sigma Xi had percentile ratings below 75, while seven and twenty per cent, respectively, were below 50.

Failure of the O. S. U. Test to measure the abilities recognized in awarding honors in the College of Agriculture and in some departments in Engineering is further indicated in the data on department honor books awarded to outstanding students in the various departments as shown in Table III. Of the mean percentile ratings definitely below 75, that is below the upper quartile, four are for departments in the College of Agriculture; Agronomy (49.1), Animal Production (49.0), Home Economics (60.2), and Veterinary Science (53.6) and the remainder in Mechanical Engineering (54.5), Law (65.7), Military Science (66.9), and Physical Education for men (48.0) and women (61.8). The five highest ratings: Economics and Sociology (93.8), Chemistry (92.2), English (91.2), Greek and Latin (90.6), and Zoloogy (89.8); are for departments in the College of Liberal Arts and Sciences. It should be observed that all of these ratings are in or very near the highest ten percentile.

It is also of interest to note that the mean percentile rating for students receiving the Gamma Sigma Epsilon award (80.8 first-year chemistry) is much lower than for the upper classmen receiving the department honor book in Chemistry (92.2), while the reverse is true for the Alpha Zeta award (59.1-agriculture freshmen) and the recipients of the honor books in Agriculture (about 49) and also for the Phi Upsilon Omicron award (82.4-home economics freshmen) and the recipients of the Home Economics department honor book (60.2). Thus, it seems that the predictive value of psychological tests for determining students who may win honors varies not only with the departments but also with the particular honor conferred by the department.

Although many students may earn high percentile ratings on psychological tests, only a very limited number of them receive

TABLE III.—RANGE AND MEAN OF PERCENTILE RATINGS ON O. S. U. TEST FOR STUDENTS ON WHOM DEPARTMENT HONOR BOOKS WERE CONFERRED

Department	Number	Range	Mean
Agronomy	15	20-85	49.1
Animal Production	16	7-90	49.0
Botany	17	59-99	82.0
Chemistry	18	80-100	92.2
Commerce	19	62-100	81.5
Economics and Sociology	16	80-100	93.8
Education			
Elementary	17	15-98	74.6
Secondary	19	28-100	75.9
Engineering			
Civil	18	24-100	82.7
Electrical	16	24-99	74.2
Mechanical	17	34-92	54.5
English	18	66-100	91.2
Geology	15	41-96	75.3
Greek and Latin	10	72-100	90.0
History	17	29-100	85.0
Home Economics	19	30-90	60.2
Law	15	11-96	65.7
Mathematics	17	39-100	82.1
Military Science	16	8-98	66.9
Modern Languages	18	28-100	81.2
Music	20	6-100	72.4
Physical Education			
Men	17	7-95	48.0
Women	19	11-98	61.8
Physics	17	6-100	77.5
Political Science	19	17-100	74.4
Philosophy and Psychology	20	23-100	85.3
Veterinary Science	15	8-100	53.6
Zoology	19	80-99	89.8

honors. This is due in part to the comparatively small number of honors awarded, but even more, perhaps, to the fact that many able students concentrate their efforts on extra-curricular activities at the expense of grades, while others are handicapped in their study by emotional difficulties and various other personality maladjustments. For many students, winning honors is a form of adjustment.

It would seem then that while there is a positive correlation between psychological test ratings and honor awards the honor recipients are not limited to students with high psychological test scores, and that many students who make high test scores fail to win honors. The percentile ratings of students winning a given honor from year to year and also for students winning different honors in various departments vary widely, in some instances even from the lowest to the highest decile.

REFERENCES

1) D. S. Hill, B. R. Buckingham, C. A. Ruckmick, and C. E. Holley. "Results of Intelligence Tests at the University of Illinois," *School and Society*, Vol. 9, 1919, p. 542.

2) E. L. Noble, and G. T. Arps. "University Students' Intelligence Ratings According to the Army Alpha Test." *School and Society*, Vol. 11, 1920, p. 233.

3) A. F. Payne, and J. D. Perry. "Intelligence Ranking of 250 City College Honor Students." *School and Society*, Vol. 42, 1935, pp. 383–4.

4) Lillian G. Portenier. "Aspects of Socio-economic Status Which Differentiate University Students." *Journal of Applied Psychology*, Vol. XXII, No. 3, June, 1938, pp. 261–270.

5) ———. "Differental University Selection." University of Wyoming Publications. Vol. VIII, No. 4, April, 1941, pp. 61–67.

BOOK REVIEWS

CARNEY LANDIS AND M. MARJORIE BOLLES. *Textbook of Abnormal Psychology.* New York: The Macmillan Company, 1950, pp. 631.

The greatest problem of modern civilization today is not atomic energy or biological warfare but is in the field of human relationships. This is one of the two fundamental ideas that the authors of this revised text in abnormal psychology have kept in mind in selecting and organizing the materials for the revised edition. The other is that the book written by them five years ago is outdated as a result of the vast amount of research that has been done in knowledge areas that are basic to understanding abnormal behavior. Their wish is, therefore, to write a revised edition that incorporates this research and also helps to enlighten the reader so that he gets a sense of the significance of human relationships out of it as well as learning the facts concerning abnormal behavior from it. Adding modern research contributions, of course, necessarily makes it possible for them to eliminate material less in line with modern tendencies. The general pattern, however, is similar in nature to that found in the first edition. The materials are presented in five sections called Orientation, Varieties of Abnormality, Explanations, Psychopathology, and Diagnosis and Therapy. Wherever possible in presenting a problem or a general topic they describe the facts known about the problem, explanations for these findings, and the psychology or psychopathology behind it. Each chapter is followed by references for further reading. The book contains a fairly comprehensive glossary and an index as well as a listing of acknowledgements of many quotations, figures, and tables used.

The kind of material that used to compose a full-sized book in abnormal psychology is here included in one section called "Psychopathology." Treated here are all the disorders of behavior: sensation, perception, action, emotions, volition, impairment of intellect, aphasia and amnesia. The section on "Varieties of Abnormality" includes the usual listing, but gives much more attention to facts recorded in recent literature about physical illness, physical handicaps, and personality adjustments of people with physical handicaps. The section on "Explanations" includes a discussion of heredity and environment, development, internal environment, and the brain. The chapter on "Psychotherapy" includes a brief consideration of the most frequently used projective techniques. The section on "Diagnosis and Therapy" includes chapters on the law, psychotherapy, educational guidance and counseling, and mental hygiene.

The chapter on "Culture and Environment" is relatively adequate and includes a consideration of psychopathology in different cultures; national, racial, and religious groups; war and economic depression; education; economic status, rural and urban rates. For some reason there are some omissions of well-known ecological studies of behavior. The viewpoint in this volume as in the first is eclectic and a conscious attempt is made to build up in the reader and the student who will use it as a text a respect for facts and figures to weight evidence with and also conscious is the attempt to be open-minded but preserve the skepticism of the scientist.

The differences between psychosis and neurosis are dealt with in line with modern knowledge; namely, that people who are neurotic more chronically are not the kind of people who get psychoses. Alcoholics Anonymous is given a relatively uncritical treatment here as in many other books recently written. Psychoanalysis, hypnoanalysis, and narcoanalysis are considered by the authors as methods of getting around the resistance of the person so he relates material which cannot readily be brought to consciousness. Group therapy, the method of the AA, is considered as forms of supportive therapy.

All in all, this is a very comprehensive text incorporating materials and contributions from recent developments of both the biological and social sciences with due emphasis to research findings. The material is all relatively well organized for teaching purposes and there is enough here to begin the consideration of other problems that are not treated at length. There is enough here to serve the purpose of an adequate course in abnormal psychology or psychopathology. Generally speaking, the people who have liked the first volume for teaching purposes will find this volume at least as good and in many respects better.

H. MELTZER

Psychological Service Center
St. Louis, Missouri

ROY M. DORCUS AND MARGARET HUBBARD JONES. *Handbook of Employee Selection.* New York: McGraw-Hill Book Co., 1950, pp. 349. $4.50.

The literature on employee selection is extensive but the quality of publications shows wide variability. In an effort to bring together minimal information on studies reporting results of selection methods the compilers of this volume examined over twenty-one hundred references. To be included in the final list of four hundred twenty-six references abstracted a report had to meet five criteria: the type of employee specified; the number of subjects given; tests used named or described; the criterion of job proficiency explicitly stated; and the actual (quantitative) results obtained reported. It is an interesting commentary on research in this field that barely one-fifth of the reports examined could meet these minimal criteria.

For each study in the final collection the abstract includes the bibliographic reference, the number and description of subjects, tests used, the job proficiency criteria, the validity data (percentages, correlation, distribution, etc.), and, where given, data on reliability. The entries are arranged chronologically and include studies between 1906 and 1949. There are indexes by job, test, and author. The job index lists two hundred eleven different jobs including teacher (with fifty abstracts), school principal, library assistant, and dean of women.

This volume should prove useful for personnel administrators and for vocational counselors.

C. M. LOUTTIT

University of Illinois

STANLEY D. PORTEUS. *The Porteus Maze Test and Intelligence.* Palo Alto, Calif.: Pacific Books, 1950, pp. 194.

Over a generation, psychological testers have passed from dependence upon a few tests, chief among them the Binet, to the active use of a large number of tools each of which has special merit for some purposes. Among the latter is the Porteus Maze Test. Porteus' new volume assembles his knowledge, experiences, and reflections relating to the Maze Test. The book is more than a manual for the test, although it includes instructions for testing, scoring guides, and similar clinical helps.

Porteus discusses use of the test in clinical diagnosis, and the implications of changes in performance following lobotomy. He presents a qualitative score which has been found to differentiate delinquents from control groups. Both of these will have interest for some practicing psychologists, although it appears that the test is best treated as an opportunity for observing the subject rather than by any objective scoring of the product. Two chapters on differences among racial groups collect much material previously published in scattered sources, and express the author's view that such differences on the Maze reflect constitutional factors.

The book is most of all a personal history, recounting the long *period when Porteus' contributions were, in his opinion,* overshadowed by the Stanford-Binet. This retelling of our tribal conflicts is not without interest, but Porteus' anecdotes, quotations from early sources, and retracing of the path by which we have modified our interpretation of the Binet test seem lacking in present-day relevance. His advocacy of the Maze Test as a clinical tool is well-warranted, but he seems to distinguish inadequately between what the Maze measures and 'planning' as a general trait. The strongly-felt opinions Porteus drops are in turn intriguing and stimulating, amusing, and outrageous. The book is a lively memento of a psychological pioneer.

LEE J. CRONBACH

University of Illinois

J. E. WALLACE WALLIN. *Children with Mental and Physical Handicaps.* New York: Prentice-Hall, Inc., 1949, pp. 549.

Specialists in the field of handicapped children will recognize in Dr. Wallin's new volume, *Children with Mental and Physical Handicaps*, an amplification of his earlier volume on the education of handicapped children. Forty years of first-hand experience as director, supervisor, and instructor in the fields of education and treatment of handicapped children has made it possible for the author to write knowingly. In twenty-one chapters he covers most phases of the field with the emphasis on diagnosis and treatment. Although the book is not divided into divisions the chapters group themselves in three large divisions; namely, (1) Definition of concepts, (2) Classification problems, (3) Consideration of special types. Classifications are treated in terms of psychological characteristics, educability and socio-vocational competency, etiology and special clinical types. Types considered include cretinism, mongolism, growth anomalies related to endocrine disturbances, anomalies related to metabolic disturbances, epileptic seizures in children, encephalitis, the orthopedically disabled, cerebral palsy, poliomyelitis, microcephaly, hydrocephalus, syphilitic brain defects. Of the thirty-four illustrations included few are new; many will easily be recognized as having appeared in earlier volumes.

On page 28 of the book, in consideration of the Wechsler-Bellevue classification distribution, the span between 70–80 is omitted. As in his previous writings of this nature the author in this volume is thorough in his descriptions, critical in his attitudes, and insistent on accurate definitions. And being an old-timer it is understandable that he refers to his earlier writings to prove that what is now being said he has already said. He not only is extremely critical of [illegible] style of environmentalism but is particularly critical of Bernardine G. Schmidt's study *Changes in Personal, Social, and Intellectual Behavior of Children Originally Classified as Feebleminded*. In the light of the facts presented by him he is justified in this criticism. His portrayal of types, while adequate for educational purposes, is not always for psychotherapeutic purposes, but the book was not intended to do everything. For consideration of personality problems with greater emphasis on sociological and cultural orientation the reader can search elsewhere. However, for a factual presentation of biological and psychological facts considered this is an excellent volume for specialists in the field.

H. Meltzer

Psychological Service Center
St. Louis, Missouri